ASSOCIATES NAL TODAY

NEW SERIES VOL. 2, NOS. 3/4 OCTOBER/DECEMBER 1977

COMPUTERIZED LITERATURE DATA FILES IN AGRICULTURE: A SYMPOSIUM IN HONOR OF THE ONE MILLIONTH CITATION IN THE AGRICOLA (FORMERLY CAIN) DATA BASE

The Associates of the National Agricultural Library, Inc. Beltsville, MD 20705

Associates NAL Today is mailed free to members of The Associates NAL, Inc. In addition to underwriting the cost of this quarterly journal, The Associates publish occasional monographs, sponsor symposiums, and initiate projects designed to supplement programs of the National Agricultural Library. Readers are cordially invited to join The Associates NAL, Inc., to receive these publications and participate in these activities. Categories of membership: Personal, $10; Business, $500; Institutional, $100; NAL staff members (nonvoting), $5. Please make your check payable to The Associates NAL, Inc., 10301 Baltimore Boulevard, Beltsville, MD 20705.

Single copies of *Associates NAL Today* are available at $5 per copy.

MANUSCRIPTS

1. One copy should be sent to
   Leila Moran
   National Agricultural Library
   Beltsville, MD 20705

2. Submission of illustrations is encouraged; authors are responsible for obtaining copyright permission.

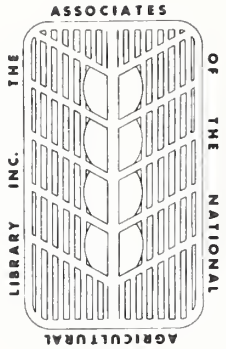
3. Manuscripts should be typed, double-space.

Cover photo: Outside view, National Agricultural Library (Courtesy, National Agricultural Library)

Associates NAL Today

New Series vol. 2, nos. 3/4
October/December 1977

CONTENTS



FOREWORD

by Richard A. Farley, Acting Deputy Director
Technical Information Systems
of the
U.S.D.A. Science and Education Administration

In 1974 when I was a candidate for the directorship of the National Agricultural Library I came away from my first interview with a pervasive thought--that NAL has an automated information system that really works. This is the fascinating story of the development of the NAL data base.

Born in a time of restricted budgets it stands today, at its millionth citation celebration, as an example of what can be done when cooperation occurs between the public and private sectors. It serves also to illustrate what dedicated public servants can do when they set their minds to producing a practical and useful product.

A word of appreciation is due The Associates of the National Agricultural Library, Inc., for sponsoring the symposium and bringing together the papers in this publication. In doing so they continue their useful function of assisting NAL in the dissemination of agricultural information to the people of the United States and the world.

INTRODUCTION

From the early cuneiform inscriptions on clay tablets to the immense informational resources of today, civilization has been faced with the growing problem of controlling information. Around the middle of the 16th century, Conrad Gesner, the father of bibliography, attempted to bring a measure of control to the literature of his time by preparing a four-volume treatise, *Biblioteca Universalis* (1545-1555), an index to Greek, Latin, and Hebrew writers.

By the end of the 19th century, it was apparent to some observers in America that the application of technology might have a positive role to play in the control and dissemination of informational data and literature. In 1886, computer generated information in the federal government began under the direction of Herman Hollerith, a former employee of the Bureau of Census, who experimented with cards and hard-wired card processing machines for census data. In spite of this early success with computer-generated information, most bibliographic and/or indexing activities involving printed literature continued to be handled manually. In 1945, Vannevar Bush in his historic essay in the *Atlantic Monthly* entitled, "As We May Think" sounded the alarm on the outmoded methods for controlling information in the printed literature.

In spite of the warnings by Bush, the 1950's saw the continuation and burgeoning of indexes which were still being created, maintained and exploited manually, thereby resulting in excesses in time and labor on the one hand and limited coverage on the other. The usual response to the increase in the volume of printed literature was the demand for an increase in budget and staff. The information explosion continued to worsen. In 1968, the world-wide output of printed material included an estimated 400,000 books, 200,000 periodicals, and 200,000 technical reports.

In particular, the problem of controlling and disseminating agriculturally related literature was no exception to the overall problem. The accumulative effect of this process is vividly illustrated by the U.S. Department of Agriculture's National Agricultural Library which by 1970 had already 1.5 million volumes of printed material and an aggressive collection-building policy for acquiring additional material. How to improve access to the expanding flow of printed information was a critical problem.

In 1970, the Library's manual reproduction of the *Bibliography of Agriculture* was replaced by an automated program that included the development of a more timely magnetic tape record of all significant agriculturally related materials acquired by the Library (NAL). Since then, operations and techniques involved in the procurement, recording, maintenance, and handling of library materials at NAL have been profoundly influenced by the advent of both the information explosion and the introduction of automated information retrieval. With the steady increase in the amount of computer-based information and the expanding application of the computer in the storage, manipulation, and display of user-oriented bibliographic data at NAL, the user has been able to obtain relevant literature in a more timely manner. AGRICOLA (AGRICultural On Line Access) is a family of data bases consisting of indexes to general agriculture, food and nutrition, and agricultural economics information. In addition, through commercial on-line services, NAL also provides access to more than 50 other data bases. Improving access to research data will remain a continuing goal and challenge for information managers at the national and international levels.

The symposium in honor of the "One Millionth Citation in the AGRICOLA (Formerly CAIN) Data Base," held on October 21, 1977, and sponsored by The Associates of the National Agricultural Library, Inc., Beltsville, Maryland, was a fitting tribute to present achievements in improving the control and access to agricultural information and also the future need for greater cooperation in all phases of the automated process among the many bibliographic data bases. This issue of the *Associates NAL Today* includes nine outstanding papers presented at the one day meeting by representatives from government and private enterprise, including land-grant college libraries as well as an information center in Holland. The papers were diverse in nature, interesting, and wide-ranging in opinion with several speakers calling for more cooperation and resource sharing on a national as well as an international basis.

Editors,
Alan Fusonie
Leila Moran

DEVELOPMENT OF CAIN HUMAN INTEREST

by

Vern J. Van Dyke*

This month the one millionth record will be added to the CAIN data base of the National Agricultural Library. To those of us who have striven to reach this goal, this is a very impressive milestone. This did not happen overnight. It has been a process of evolution, sometimes funny, sometimes disappointing, with always a lot of hard work and cooperative effort. Under a manual system, people can be individualistic; under a computer system, there must be standards, and standards must be followed.

Some people believe that all one has to do is "Press a button." Well, NAL has pressed many buttons--including the panic button.

The first button pushed to reach this goal of one million records took place way back in the early 1960's when a group of professionals from many fields joined together in what was called Task Force ABLE to perform a management and systems study of NAL to determine whether computer technology could effectively assist the library in servicing scientists and researchers. ABLE was the acronym for Agricultural/Biological Literature Exploitation. Recommendations included the use of computers and suggestions for priorities.

With this study completed, efforts were set in motion which have brought us to this milestone. The Agricultural/Biological Vocabulary was automated only to discover later that a structured vocabulary was too restrictive for widespread usage in a cataloging and indexing system. The personal author and subject indexes were automated only to discover that users asked for more and better indexes, catalog cards, book catalogs, data bases for retrieval purposes, online input, and selective output. And still the demand is heard for more and better ways of utilizing our data base. When will it end? Who knows? It is a great challenge calling not only for highly evolved computerized technology but also for broader and more flexible skills on the part of librarians themselves. As the computer systems grow, so must also the people using them.

In 1965 NAL's published data were available mostly in libraries and may have reached only a limited audience. Today CAIN is available to the scientist and researcher in his office, laboratory, or in the field. The data base on magnetic tapes is sent monthly to Sweden, Norway, Mexico, South America, Netherlands, Canada, and Australia. In addition, three commercial sources offer CAIN in an online searchable capacity--two of which are also available in Europe.

There is one thing that should be clarified. CAIN, which stands for CAtaloging and INdexing, is the NAL data base, but CAIN also is a computer system which can and does handle other data bases.

All early efforts are subject to trial and error. We at NAL have learned a lot from early problems and are still learning. I would like to share with you some of our more humorous situations.

In September 1965, a contract was awarded to write programs for the Pesticides Information Center, the contract to be completed by September 1966. When I started working at NAL in September 1966, the first program had not yet been written, the contractor's reason being "IBM is having problems with their compiler and we are not going to debug it for them." Imagine our surprise when at a meeting in February 1967, the contractor announced they were writing the system for a CDC 3200; we had an IBM 360. When we objected, the contractor's legal representative threw a copy of the contract on the table and said, "Show me where it says what machine we are to write the program for." Lo and behold, it was not mentioned. So they wrote programs in COBOL for a machine we did not have.

I think that Nancy Ayer, Carl Cantrell, and I trained at least four new project leaders for the contractor. These project leaders would start working on the system and after a short time throw up their hands and quit.

In November 1967, the first *Pesticides Documentation Bulletin* was produced using the contractor's computer. Charles Bebee, who was then Chief of the Pesticides Information Center, and his staff shed a lot of blood, sweat, and tears to get the first two issues out. And again imagine our frustration when the January 1968 issue could not be produced because the contractor's project leader had decided to make some changes. He did not, however, bother to duplicate the programs, kept no listings and, after the changes, could not get the programs to run.

We at NAL finally got hold of the programs and converted them to the IBM 360. But don't tell Nancy, Carl, and me that COBOL programs run on different machines with only minor changes.

At the same time we were doing a systems study for what we believed to be a complete library system. It was called "Document Locator and Control System" (DLCS) and in February 1969, Nancy Ayer prepared a Functional Schematic for that system. Although yellowing with age, it is still in fairly good condition in my office but it was never actually used because of budget restrictions.

About that time John Sherrod became Director of the Library and Abe Lebowitz became Deputy Director for Resource Development. They had different ideas of what the Library should be doing. So, after discussing and deciding what was needed, we felt that the PIC system, as it became known, could be modified and expanded to do the job. In many meetings with Abe Lebowitz, Jeanne Holmes, John Forbes, Carol Johnson, and many more people from Resource Development CAIN was born.

And so after a lot more sweat, blood, and tears CAIN made its first production run for the January book catalog and weekly catalog cards. And, of course, we got such questions as: "What's this?" "Why does this look like this?" To our answer "that's what you asked for," their reply was "that's not what we meant;" so changes were made.

Then Abe Lebowitz decided to accept a position in Israel and Sam Waters became Deputy Director for Resource Development--and, of course, he had different

*Vern J. Van Dyke, Systems Analyst at the National Agricultural Library

ideas of what was needed by NAL in the way of automation. So, after many more meetings, some of us becoming quite emotional with table pounding, etc., because he wanted to change OUR system, we did finally negotiate changes. We did them his way.

Jeanne Holmes, I'm sure, will talk about data input to the CAIN system but I do want to mention several frustrating incidents. Our input at the start was in the format of cards that were punched by a contractor. In one case if we received cards with only a ninety percent error rate, we considered the contractor had done a good job; when complete batches were returned to be repunched that contractor didn't last too long. Another keypunch contractor was in Kentucky. Although their error rate was very low, delivery was a problem. We did somehow manage to run the system for two or three years by making changes here and patching there and writing new programs to do things that were never thought of at the beginning.

In 1972 we made some major changes so it could be the "perfect" system for NAL. Comments we were receiving from the users were taken into account and corrections or changes made when possible. We were now going to sit back and watch our "perfect" system run with no complaint from the Catalog Section, Indexing Section, or anyone else. We were right--no complaints, nothing wrong--until we ran it the first time.

The CAIN system began to come to the attention of USDA agencies wanting its flexibility and capability for their own data. Separate data bases were subsequently created by Agricultural Economics and Food and Nutrition Information Centers and both of these data bases are now offered with NAL's CAIN data base.

Some organizations retain their own data in the CAIN format separately as they have different uses. Examples of these are the Herbicide data of the USDA Agricultural Research Service, the International Tree Disease Register of the U.S. Forest Service, the *List of Available Publications* of the USDA Office of Communication, and the several data bases of the Automated Data Systems, USDA Agricultural Research Service.

Other new data bases that will be coming under the AGRICOLA umbrella in CAIN format may include the following: the Arid Lands Information and the Water Resources data, both from the University of Arizona; SAHEL data from the University of California at Riverside; Brucellosis data from APHIS (Animal and Plant Health Inspection Service), and another data base from APHIS called EPIC (Emergency Programs Information Center). And still the list continues to grow.

Now there has to be something good about a system which can handle this variety of information. Such aspects might include the following:

1. CAIN accepts data in MARC type format;

2. Since 1973 it has been possible to enter data online through a minicomputer for editing and processing before remote job-entry to the larger computer file;

3. CAIN produces output in upper/and lower case which is natural to and desired by librarians and other users;

4. Since 1974 we have been able to earmark records from the OCLC (Ohio College Library Center) system for transmittal to CAIN via magnetic tape;

5. Extensive output formats include catalog cards, book catalogs, bibliographies, etc., in either upper/lower print train or formatted for acceptance by the Linotron;

6. Commercial firms accept the CAIN format and make it available for selected online retrieval.

Having been involved with the computer side of CAIN, analysis and programming, from the beginning until the end of 1974, I know the unreasonable demands with which we were faced. As a member of the Director's staff, one of my duties is getting new data bases into AGRICOLA through CAIN. To this end, some modifications are again being made.

NAL does not contend that CAIN is the best or ultimate system but it has been and will probably continue to be useful for some time to come. It serves many purposes and no one can ask more of a system. It has enabled us to reach the one millionth citation. Will it reach the second million? I believe it will.

DEVELOPMENT OF NAL DATA BASES

Input by Resource Development

by

Jeanne M. Holmes*

The main title of the topic assigned to me--Development of NAL Data Bases--is sufficiently general that it suits my purposes beautifully. It's very obvious, however, from the subtitle, that the planners of this symposium intended for me to talk primarily about Resource Development input to the CAIN data base, NAL's machine-readable CAtaloging and INdexing records which make up the major portion of AGRICOLA, the AGRICultural On-Line Access service.

And, indeed, I will address that topic. But, exercising the privileges of sex and seniority, I am going to broaden the scope of my remarks in order to give, I think, a better perspective of the Library's efforts to provide access to the world's agricultural literature.

Vern Van Dyke referred to the establishment of Task Force ABLE in 1962 as "the first button pushed to reach this goal of one million records" which we commemorate today. I have no quarrel with Vern about that statement. I would merely shift the emphasis a bit to be sure that everyone understands that "*this* goal of one million records" is only one noteworthy achievement in a continuum. It has been preceded by and, I am confident, will be succeeded by other events of bibliographic significance in the NAL story.

For this presentation, the NAL story, like the story of man, might be divided into two parts: B.C. and A.D. "Before Computers" would do nicely as the full name for the B.C. period. I hesitate, though, to spell out A.D.; Automation for "A" is a natural assumption, but the "D" might generate the thought of "deity" for some and of "demon" for others. Perhaps we could settle for "Automation Days," in which "days" is defined as a period of power or influence.

BEFORE COMPUTERS (B.C.)

Be that as it may, as I see it, the beginning of the evolutionary process which today has peaked--again--was in *1862*, rather than 1962. The earlier date, of course, marked the establishment of the U.S. Department of Agriculture and its Library.

The Agricultural Division of the Patent Office in the Department of Interior was the foundation for the new Department. And the Division's collection of 1,000 publications on agricultural statistics formed the nucleus for the Agriculture Department's library collection. From this small beginning, the collection has grown steadily in numbers and in subject scope. In 1962, the millionth accessioned volume was added. By that time, the collection scope had been expanded to cover agriculture in its broadest sense, and acquisitions included publications from all over the world.

It is obvious from archival records that it was never the intent of Library administrators to acquire books and just put them on the shelves to collect dust. They were meant to be used by those who needed them. Access to information was of primary importance from the earliest days of the Library. This access has been provided through bibliographic tools, such as catalogs, indexes, and bibliographies.

Catalogs

Hard facts about the development of data bases in the 1860's are difficult to come by. It is known, however, that by 1871 a catalog was in preparation. The ambitious plan called for author and subject access as well as the inclusion of short abstracts. Ambitious the plan may have been; successful it was not! In 1880, money was sought in the budget so that the collection could be *properly* cataloged.

Eight handwritten volumes, completed in 1887, contained the *Class List of Works in the Library of the United States Department of Agriculture*. This, incidentally, reflects the earliest classification scheme and is the first book catalog of the Library of which I am aware.

The same year that the *Class List* was completed, 1887, the Library's catalog was firmly and publicly declared to be totally inadequate. Thus has been the case in the intervening 90 years--and thus it will probably ever be!

Undaunted and always mindful of the need to mend their cataloging ways, subsequent librarians devised and revised another classification scheme until finally the Library of Congress system was adopted in 1966. As new cataloging rules and standards have been developed by the profession, they too have been adopted. Policies have been revised, the better to meet the identified needs of the agricultural community. All of these changes have been reflected in the Library's catalogs, from the earliest classified list to the current catalog. The catalogs also reflect advances in technology, from handwriting through mimeographing and photoreproduction to computer production.

Indexes

The recorded history of NAL indexes to agricultural literature is somewhat, but not considerably, shorter than the lifespan of the catalogs. That stands to reason. First, one acquires and records the collection; then one can more readily delve into its content.

Indexing Section apparently did not experience the frustrations of the cataloging process. Or, at least, such frustrations seem not to have been widely publicized. By 1900, card indexes to the publications of the U.S. Department of Agriculture were being compiled, including one covering all articles in the USDA *Yearbooks* from 1894 through 1898.

About the same time, 1900, libraries in the individual bureaus of the Department became noticeably active.

---

*Jeanne M. Holmes is Deputy Director for Resource Development at the National Agricultural Library.

And many of them engaged in the compilation of recurring bibliographies of the current literature in the bureaus' special fields of interest. Then, in 1942, the bureau libraries were consolidated into the Department Library, and such bibliographies as *Entomology Current Literature* and *Plant Science Literature* became sections of the newly created *Bibliography of Agriculture*.

Within 18 months of its inception, subjects were added which had not been covered by the bureau libraries; a classification system was devised which integrated the parts of the *Bibliography* more closely; the format was changed; and author and subject indexes were included. Changes that have been made in the ensuing years have been either planned as improvements or they have been the inescapable result of circumstances, both economic and technological. Some of the changes, particularly in recent years, have been controversial--so controversial, in fact, that I'm going to take the coward's way out and leave that topic for another occasion and, preferably, another person. It should be noted, though, that, as originally conceived, the *Bibliography of Agriculture* was to be comprehensive. And this is the goal toward which the Library continues to strive, admittedly with only partial success. Production techniques employed for the *Bibliography* have closely resembled those associated with the catalogs, with one major exception. I've found no evidence of a handwritten edition.

Bibliographies

I mentioned earlier that the Library has traditionally provided access to its collection through products such as catalogs, indexes, and bibliographies. The meaning of "catalogs" and "indexes" is usually clear to everyone. But a definition of "bibliographies" in the context of this paper is in order, I believe, especially in view of the title of the *Bibliography of Agriculture*.

By "bibliographies," I mean lists of publications, generally nonrecurring, that are limited to special subject fields but not limited to the current literature in those fields. Such bibliographies are now prepared by the Reference Division, not a unit of Resource Development. But a mention of them is needed to complete the story of both the past and the present. A few highlights should suffice.

Evidence exists that as early as 1894 the Library was engaged in the compilation of special subject bibliographies. Some of the topics addressed in the late 19th and early 20th centuries were forestry, chemistry, and botany, as well as more specific subjects such as irrigation and land drainage.

In 1942, the *Library List* series of special bibliographies was begun. It was continued even after the initiation in 1943 of the now defunct printed departmental series, *Bibliographical Bulletin*. And with varying degrees of frequency, special bibliographies have continued to be prepared to this day.

AUTOMATION DAYS (A.D.)

It must be very apparent by now that I view the work of today's Resource Development staff as differing very little from that of the librarians of the fledgling Library of the United States Department of Agriculture. They acquired, to the best of their ability, agricultural publications from all over the world. They organized those publications to make them available and analyzed their contents. Those librarians also created bibliographic products to announce the availability of the publications. And, of greatest meaning to this symposium, each generation availed itself of the advantages of advances in professional thinking and technology, even if, sometimes, with fear and trepidation.

This is the heritage of CAIN and AGRICOLA. Resource Development is but one link in the chain. If there were time, the work of each predecessor unit should receive equal attention. But the CAIN data base is in today's spotlight, and Resource Development is CAIN's producer.

First, a few words about Resource Development itself. For the uninitiated, Resource Development is simply another name for the functional grouping which, in another library, would probably be called Technical Services. In fact, in its 1963-to-1970 incarnation, the present RD functions, with the exception of indexing, were the responsibility of an NAL Assistant Director for Technical Services.

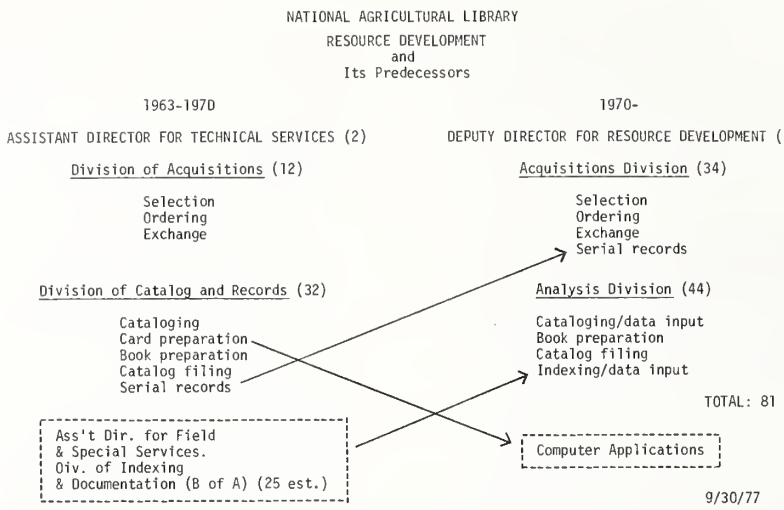
NATIONAL AGRICULTURAL LIBRARY
Resource Development

9/30/77

	Librn/TIS	Lib Tech	Clk-Typ	Total
Procurement	4	15	1	20
Selection	6	5	1	12
Acquisitions Chief	1		1	2
TOTAL				34
Cataloging	11	11	1	23
Indexing	13	7		20
Analysis Chief	1			1
TOTAL				44
Resource Development	2		1	3
	38	38	5	81

Resource Development is responsible for acquiring and organizing library materials. It is the INPUT unit of the Library--input to the collection and to the bibliographic data base.

Organizationally, RD is made up of two divisions: Acquisitions and Analysis. Each division has two sections: Selection and Procurement in Acquisitions; Cataloging and Indexing in Analysis. The total RD staff as of September 1977 numbers 81 (38 professional; 43 paraprofessional and other support staff). All units contribute directly or indirectly to the development of the data base.



Cataloging

Selection and ordering of new publications are the responsibility of the Selection Section, as well as the assignment of priorities for cataloging and processing of materials and the determination of which monographs in series will be analyzed for the data base.

Once acquired, new books and serials, logically and actually, go to the Cataloging Section for processing in the order determined by Selection's assigned priorities. Cataloging uses the Anglo-American rules for entry and description, Library of Congress classification, and Library of Congress subject headings. In addition, a broad subject category is assigned to each title cataloged. The category is used for arranging citations in the commercially published catalog and for retrieval from the data base. Also, the title of each foreign-language work is translated into English, another aid to retrievability.

The vast majority of the cataloging records are entered into the computerized CAIN file through the Ohio College Library Center (OCLC) on-line shared cataloging system, based in Columbus, Ohio. Magnetic tapes of NAL records in the OCLC format are converted to the CAIN format for internal processing. Catalog cards are created from CAIN, as well as camera-ready copy for the commercially published *National Agricultural Library Catalog*, which appears monthly.

A few cataloging records which are not within the scope of the OCLC system, such as translations of articles, are input on-line to CAIN through the minicomputer data entry and edit system which Vern Van Dyke mentioned. All updates and approvals of CAIN cataloging records are entered through that system.

About 14,000 titles are cataloged each year. They comprise about 10 percent of the CAIN data base.

Indexing

By simple arithmetic, it is easy to determine that 90 percent of the data base is made up of indexing records. All of these records are derived from previously cataloged works. Some of them are citations to non-journal literature, such as monographs in series, annual reports, and conference proceedings. Most of them, however, refer to journal articles.

Newly received issues of cataloged serials, journals and others, are recorded and processed in the Procurement Section of the Acquisitions Division. About 230,000 issues are checked in each year, representing approximately 20,000 serial titles. Issues of about 6,000 of these previously cataloged serials are routed regularly to the Indexing Section for professional selection of articles to be included in CAIN. In addition, newly cataloged titles, serial and monographic, are reviewed by indexers as candidates for addition to CAIN. Items chosen for indexing are selected on the basis of criteria designed to best meet users' needs within the limits of the Library's resources. Each year, about 120,000 indexing records are added to the data base. They are distributed on the AGRICOLA monthly sale tape from which the *Bibliography of Agriculture* is produced by a commercial publisher.

The indexing records prepared in Resource Development contain full bibliographic descriptions, including journal title abbreviations, English translations of titles in foreign languages, keyword enrichment of meaningless or ambiguous titles, and classification by NAL's broad subject categories. Certain other data are also recorded to enhance identification and retrieval; these include such data as codes for the language of the text and the source of the document.

Citations are input to CAIN on eight CRT terminals of the minicomputer data entry system mentioned earlier. This system also supports the associated authority file of journal title abbreviations used in the citations.

Bibliographies

From the computerized file of cataloging and indexing records, special subject bibliographies are created by the Reference Division. Some of them are provided, on demand, only to the individuals who request them. Others are deemed to be of wider interest and are made available as numbers of the *Quick Bibliography Series*.

CONCLUSION

Thus the Library carries on its tradition of acquiring, organizing, and making available agricultural information. The emphasis now is on the one-time creation of a single data base from which multiple products can be derived. The computer has eliminated the need for repeated human processing of the same data for different purposes.

This phase of the continuing evolution of bibliographic control at NAL has been exciting and challenging. The next one will be even more so, I am sure.

Automated online retrieval system (Courtesy, National Agricultural Library)

A SURVEY OF AGRICOLA (FORMERLY CAIN) DATA BASE

by

Anne S. Hubbard and Richard P. Caputo*

"You people have answers for which we don't yet have questions," was a frequent response heard to on-line retrieval systems in the late 60's. "It costs too much," was another. This new technology was being compared with batch processing which has been characterized as having either not enough or too much information, or with the manual search, which although it produced results, also consumed inordinate amounts of search time.

As recently as 1969, the only on-line data bases available were a few private files loaded into retrieval systems and use of these files was usually limited to the "data gathering" organization such as government agencies. The average cost per search during this time frame was $50-$1,000, which is very high by today's standards.

The 1970's brought about very dynamic and positive changes in information retrieval, as shown in Figure 1. The average cost per search today is $5-$20, with an anticipated further cost reduction by the end of this decade. Worldwide, there are now some 30-40 on-line retrieval services offering in excess of 80 data bases comprising nearly 25 million references. Users now include government agencies, industrial and academic institutions, along with an increasing number of individual professionals. In 1977 these individuals will perform 1.5 million searches and print 45 million references off-line.

In 1973 the National Agricultural Library, which is the third largest federal library in the United States, recognized the potential of on-line searching and decided the time was right to load the first major agricultural data base. Lockheed Information Systems' DIALOG(R) was chosen. Some key early dates are:

March 28, 1973	National Agricultural Library issues Request For Proposal (RFP)
April 30, 1973	Proposals due
May 16, 1973	Roger Summit and Robert Donati of Lockheed Information Systems conduct benchmark demonstration for National Agricultural Library representatives.
June 12, 1973	National Agricultural Library announces award to Lockheed
September 1973	National Agricultural Library/ CAIN (AGRICOLA) ON-LINE and National Agricultural Library personnel trained by Robert Donati and Richard Caputo

In addition to this contract, which was primarily intended to provide in-house use of the National Agricultural Library file, the National Agricultural Library made an important policy decision: The tapes would be distributed at the cost of reproduction to any service so requesting. In this way the National Agricultural Library could achieve the widest exposure of the resource they had developed at the lowest possible cost to the government.

October 1977 marks another major milestone in the chronology of AGRICOLA--the one millionth citation will be added to the on-line system. On this occasion it is appropriate to mention some of the milestones leading to that achievement; there are many.

The National Agricultural Library was a pioneer in emphasizing machine readable records. In 1970, the Library ceased publication of the *Bibliography of Agriculture* through the Government Printing Office and concentrated on the production of the CAIN (CAtaloging and INdexing) tapes--the complete bibliographic record of the Library on a current basis in computer readable form. From the tapes came the printed book catalog, a *Bibliography of Agriculture*, a Selective Dissemination of Information (SDI) service, and within a short time on-line access as well.

Cooperation has led to yet another pioneering achievement--the production of a data base including a variety of information from different sources. First, within the Library itself, the cataloging and indexing records were put on one tape with format compatibility, then the Food and Nutrition Information Center records were included, followed by the Agricultural Economics records. More recently, this exemplary cooperation has been extended internationally with Canadian records becoming part of AGRICOLA and AGRICOLA records being contributed to AGRIS, the newly emergent international file of agricultural information.

At an early date the National Agricultural Library also recognized that it could extend service beyond simply providing the data and began extensive efforts to provide an explanation of the information available as well as to demonstrate the various means of access and to provide in-depth training on the data base to a substantial number of users. This is a service that is only now being provided by other data base suppliers and is an achievement in which the National Agricultural Library can take pride.

*Anne S. Hubbard and Richard P. Caputo, Lockheed Information Systems.

FIGURE 1

RETRIEVAL SERVICES POTENTIAL

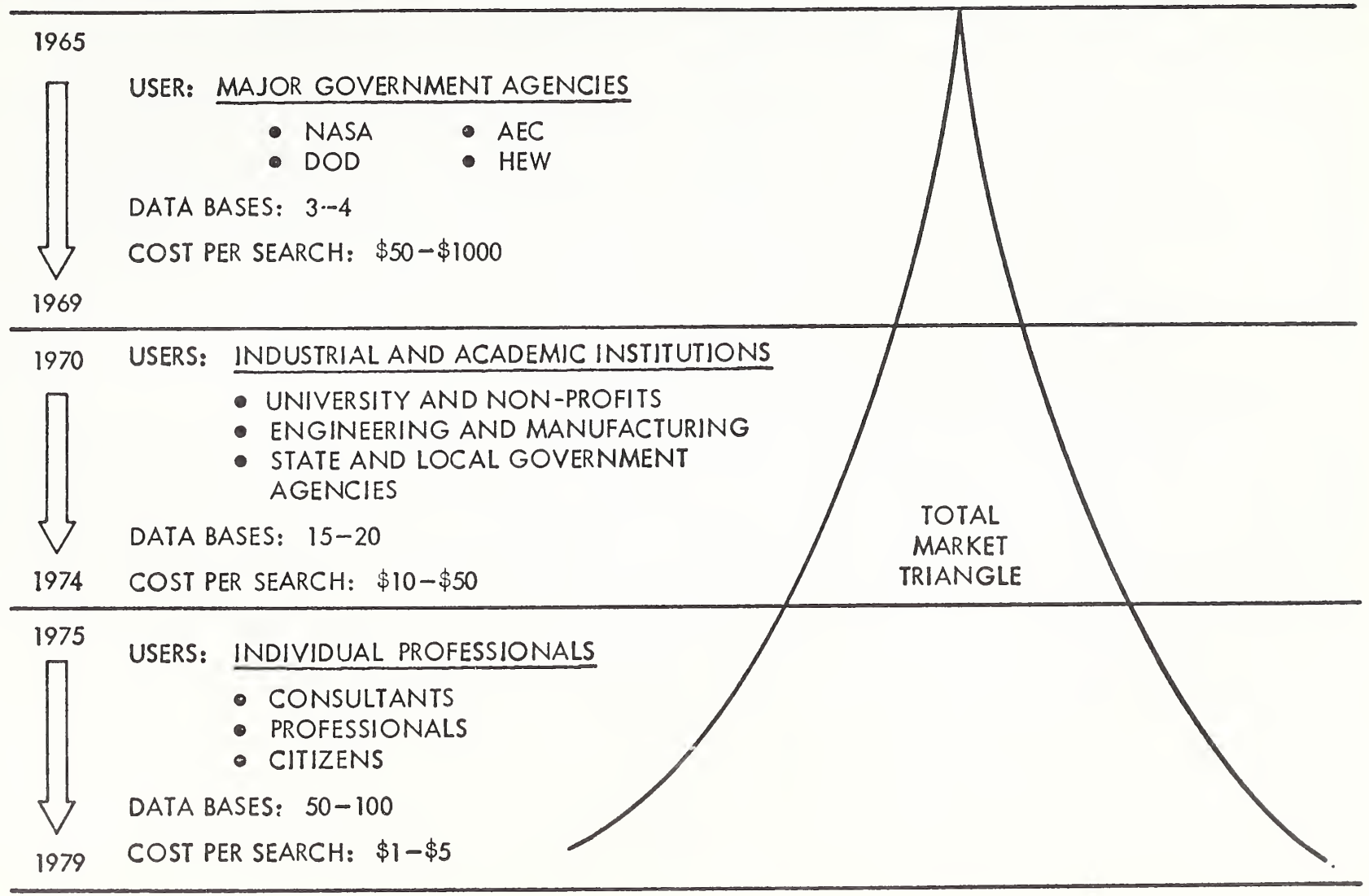


FIGURE 2. AGRICOLA SUBJECT COVERAGE

	AGRICULTURAL ECONOMICS AND ADMINISTRATION	AGRICULTURAL ENGINEERING	AGRICULTURAL PRODUCTS	ANIMAL SCIENCE AND INDUSTRY	CONSUMER PROTECTION	CHEMISTRY AND TECHNOLOGY	ENVIRONMENTAL POLLUTION	ENTOMOLOGY	FOOD AND HUMAN NUTRITION	FORESTRY	GENERAL AGRICULTURE	HOME ECONOMICS	INFORMATION SCIENCE	LIFE SCIENCES	NATURAL RESOURCES GENERAL	NATURAL RESOURCES MANAGEMENT	PLANT SCIENCE	PHYSICAL SCIENCES AND MATHEMATICS	RURAL SOCIOLOGY, SOCIAL SCIENCES, HUMANITIES	SOILS AND FERTILIZERS	VETERINARY MEDICINE
AGRICOLA	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
AIM/ARM					X							X									
AMERICA HISTORY & LIFE	X																		X		
APTIC	X		X	X			X		X								X			X	
BIOSIS						X	X	X			X			X			X	X			X
CA CONDENSATES						X	X		X						X			X		X	
CAB ABSTRACTS	X	X	X	X	X	X	X	X	X	X	X		X	X	X	X	X	X	X	X	X
COMPENDEX		X				X			X		X										
COMPREHENSIVE DISSERTATION ABST.		X		X		X		X			X		X	X	X		X	X	X		
CRIS	X		X		X		X		X	X	X	X		X	X		X				
ENVIROBIB			X			X	X		X	X				X		X				X	
ENVIROLINE			X			X	X		X	X				X		X				X	
FSTA		X				X			X		X	X		X		X			X	X	
NTIS	X				X	X	X				X		X	X	X			X	X		
POLLUTION ABSTRACTS			X				X	X		X					X		X			X	
PREDICAST PTS	X		X			X				X	X	X			X	X				X	
SCISEARCH (R)	X			X	X	X	X	X	X		X	X	X	X	X		X	X		X	X

The AGRICOLA data base provided comprehensive coverage of world-wide journal and monographic literature and related subjects. AGRICOLA provides information in the following multidisciplinary areas:

Agricultural Economics and Administration
Agricultural Engineering
Agricultural Products
Animal Science and Industry
Consumer Protection
Chemistry and Technology
Environmental Pollution
Entomology
Food and Human Nutrition
Forestry
General Agriculture
Home Economics
Information Science
Life Sciences
Natural Resources (General)
Natural Resources Management
Plant Sciences and Mathematics
Rural Sociology & Social Sciences & Humanities
Soil and Fertilizers
Veterinary Medicine

In addition to the topics listed above, AGRICOLA covers collateral or related items, such as land use, family migrations, labor, political movements, and many others.

In 1977, some 50,000 searches were conducted on-line and approximately two million items were printed off-line as a result of these searches. In other words, an equivalent of the entire file will effectively be republished twice during the year by on-line services in direct response to queries posed.

Supportive Information

Today's user of on-line retrieval searches an average of 2.3 data bases per topic. In addition to the broad coverage offered by AGRICOLA, the user is provided several supportive data bases. Figure 2 breaks down the agricultural sciences into broad topical groups and shows relative coverage by other related data bases available within DIALOG. This breakdown is useful in piloting a user to the specific data bases covering a particular topic. Although there is some duplication of coverage, each of the data bases treats the material somewhat differently with regard to indexing, abstracting, and bibliographic material provided.

To continue to define the AGRICOLA user group as just the agricultural research community is a gross understatement. Because of the broad spectrum of subjects covered in AGRICOLA, the user community is equally as broad. I personally recall a situation that occurred early this year at a large electric utility in which the requester was specifically interested in studies done on a particular type of submersible transformer. He recalled this research as having been done two or three years ago, but could not remember the author or publication. We constructed the search profile with this limited information and employed a DIALOG feature called SEARCH-SAVE(TM), a very time-saving system convenience, allowing the experienced user to build a search, save it within the system, change data bases, and execute the search in the new data base. The actual re-execution of the search is done automatically and quickly, therefore being very cost effective. In short, our requester was amazed to find three citations in AGRICOLA and two in Commonwealth Agricultural Bureaux. These were the only references found in all the data bases we searched. It illustrated the importance of AGRICOLA as a prime use data base.

The one millionth record in AGRICOLA is an impressive milestone and it is a pleasure to be part of the acknowledgement it richly deserves. It is an impressive chapter in a century old tradition of providing agricultural information. It is an achievement that would certainly astound the founders of the Department of Agriculture and the Library but, with equal certainty, it would be recognizable as exactly what was intended in establishing a library to serve the agricultural community in the United States and, indeed, the entire world.

Future Hopes and Plans

In the future, increased systems use can be correlated to expansion of the existing telecommunications networks, TYMNET and TELENET. As nodes are introduced in new locations usage increases because a local, rather than long distance, call need be made. It is hoped that both networks will continue the expansion that has been seen so far.

Systems use could further be simplified by data base suppliers providing a more uniform table of contents, with greater consistency in field names. The greater the similarity in data base structure, the less the user is required to remember and, in turn, the time required in search preparation and execution is less.

It is hoped that cooperation among the data base suppliers will continue at a more accelerated pace, with an increase in the coordination between their products. For, as systems become easier to use, they become more cost effective and total usage increases.

SDC SEARCH SERVICE--PAST, PRESENT, AND FUTURE

by

Carlos A. Cuadra and Sheryl Rosenthal*

On-line retrieval services of greater than local scope have now been in operation for over 12 years. Those years have seen dramatic changes both in the technical quality of the services available and in the sophistication of their use. And while many people believe that most of the basic capabilities that can be developed for on-line systems have already been developed, we at SDC believe that the best is yet to come.

Background of SDC Search Service

The technological roots of on-line services go back to the middle 1950's, when the concept of computer "time-sharing" was born and was implemented in the U.S. Air Defense System. Computer time-sharing involved the use of a computer to service simultaneously a number of users, interacting with the computer through a terminal. These users may have been sharing the same program and/or the same files, or they may have been working on entirely different programs and files.

SDC designed and programmed the first large-scale time-sharing systems in the middle and late 1950's, for the Air Defense Command and other military agencies. The Air Defense System computer and programs served about 150 users at terminals scattered throughout a single large building. The terminals doing the highest priority work received responses to an inquiry or to the entry of data within five seconds. Other terminals received responses within 10 or 15 seconds, the time depending upon the importance of each terminal's work. We have obviously come a long way from these early days. Nowadays, users typically receive responses to their inquiries in only two to five seconds, even though they may be located many thousands of miles from the computer.

Systems Development Corporation has been a pioneer in the field of on-line services from its inception. The name "SDC Search Service" was born in December 1972, at the time we decided to load the Chemical Abstracts Condensates data on-line for access on a world-wide basis. But, in actual fact, we had been moving toward full-scale commercial service for over a decade. By 1960, we had developed our first interactive retrieval program that operated on textual data, using what some refer to today as "full-text" searching.

By 1965, we had launched an experimental on-line service on a nationwide basis, under the sponsorship of the Department of Defense, to evaluate the potential of on-line retrieval. The users of this experimental service were 12 government agencies and private contractors in the Washington, D.C. area and several other areas of the U.S. The computer itself was located in California, at SDC headquarters. The file --150,000 bibliographic records on foreign technology --was considered in those days to be very large. Nowadays, we would consider it to be relatively small.

By 1970, the feasibility and promise of on-line services for libraries were so apparent that the National Library of Medicine decided to introduce full-scale service on their MEDLARS files, using a special version of SDC's "ORBIT®" program that NLM called "ELHILL." Some of you may have been among the early users of ELHILL, on NLM's AIM-TWX service, in 1970 and 1971, or on the MEDLINE service which succeeded it. The nationwide exposure that MEDLINE gave to on-line retrieval techniques helped to create awareness of their potential and thus contributed to the acceptance and use of commercial on-line retrieval services, such as SDC Search Service.

Since the inauguration of SDC Search Service, we have been adding about five to eight new data bases a year. AGRICOLA, originally called CAIN, was one of the first data bases to be added, in June 1973. At the current time, we are adding at least one new data base every month, to help expand the resources available to our users and to maintain our position in what has become a highly competitive retrieval services industry. The data bases available to users of SDC Search Service, as of November 1, 1977, are shown in Figure 1.

ACCOUNTANTS INDEX
AGRICOLA
API LITERATURE
API PATENTS
AMERICAN STATISTICS INDEX
BIOCODES (DICTIONARY)
BIOSIS PREVIEWS 1974-
BIOSIS PREVIEWS 1969-1973
CANADIAN BUSINESS PERIODICALS INDEX
COMPREHENSIVE DISSERTATION INDEX
CHEMCON (CA CONDENSATES) 1972-
CA CONDENSATES 1970-1971
CHEMICAL INDUSTRY NOTES
CONGRESSIONAL INFORMATION SERVICE INDEX
COMPENDEX
CONGRESSIONAL RECORD ABSTRACTS
ENERGYLINE
ENVIROLINE
ERIC
FOOD SCIENCE & TECHNOLOGY
GEO-REF
GRANTS
INFORM
INSPEC
LIBCON/E (ENGLISH)
LIBCON/F (NON-ENGLISH)
LIBRARY & INFORMATION SCIENCE ABSTRACTS
MANAGEMENT
NTIS
OCEANIC ABSTRACTS
PAPERCHEM
PESTDOC
PETROLEUM/ENERGY NEWS
PHARMACEUTICAL NEWS INDEX
POLLUTION
PSYCHOLOGICAL ABSTRACTS
RINGDOC
SOCIETY OF AUTOMOTIVE ENGINEERING
SSIE
TITUS
TULSA
WORLD PATENT INDEX

Figure 1. Data Bases Available through SDC Search Service

*Carlos A. Cuadra and Sheryl Rosenthal, Systems Development Corporation.

Types and Characteristics of On-line Retrieval Services

Some of the members of this audience are experienced on-line searchers. Others are generally familiar with on-line techniques but have no direct or extensive on-line searching experience. Still others, perhaps, have had only minimal exposure to either the concepts or the techniques of on-line searching. While it is very difficult to address all three levels of familiarity in the same discussion, a brief review of the characteristics of on-line retrieval services may help to provide something of a shared context for our remarks on the future of these services.

There are many kinds of on-line systems. Some of them are in operation in banks; others in airline companies; others in food markets and car rental agencies. Still others, of a somewhat different type, are in use in system development companies, universities, and governmental agencies to help computer programmers write their programs more efficiently by using on-line interactive techniques.

The kind of system that is of professional interest to this audience is the kind that helps a scientist, technician, administrator, librarian, information specialist, or other intermediary locate information rapidly and precisely in very large files (or "data bases") stored in a computer. Most of the data bases of interest to you are primarily bibliographic in nature. No one knows precisely how many such data bases there are, but Martha Williams[1] has recently provided an estimate, shown in Figure 2.

	No. Data Bases	% Data Bases	No. Records	% Records
U.S.	160	58%	46.3M	89%
Non-U.S.	117	42%	5.7M	11%
	277	100%	52.0M	100%

Figure 2. Number of U.S. and Non-U.S. Data Bases

About 58 percent of the data bases that met her survey criteria originate in the United States. These tend to be fairly large data bases, as witnessed by the fact that they represent 89 percent of the total number of records. AGRICOLA, of course, is one of these large data bases.

Data bases differ in a number of respects, including size, but the two most important differences are probably content and the types of source documents covered. We find it useful to think of data bases in terms of four basic types:

> Scientific-technical data bases, exemplified by AGRICOLA, Chemical Abstracts Condensates, BIOSIS Previews, Compendex, and INSPEC;
>
> Social science/humanities data bases, exemplified by ERIC, Psychological Abstracts, and Social Science Citation Index;
>
> Business data bases, exemplified by ACCOUNTANTS INDEX, Chemical Industry Notes, Pharmaceutical News Index, and World Patent Index;
>
> Multidisciplinary data bases, exemplified by NTIS (National Technical Information Service), Comprehensive Dissertations Index, SSIE (Smithsonian Science Information Exchange), and Congressional Information Service Index.

How are these data bases used? Figure 3, a schematic representation of a typical on-line retrieval system, shows users connected to the system by means of a terminal. The terminal is coupled through a telephone into some sort of communications network which puts the user in contact with a large time-sharing computer. This computer provides access to one or more bibliographic files stored on magnetic disks and users of the system are able to type in search inquiries and receive immediate responses in the form of citations or abstracts of current literature. In the case of lengthy bibliographies or large numbers of abstracts, users may prefer to have the responses printed off-line, on a high-speed printer. These off-line printouts are then mailed to the user.

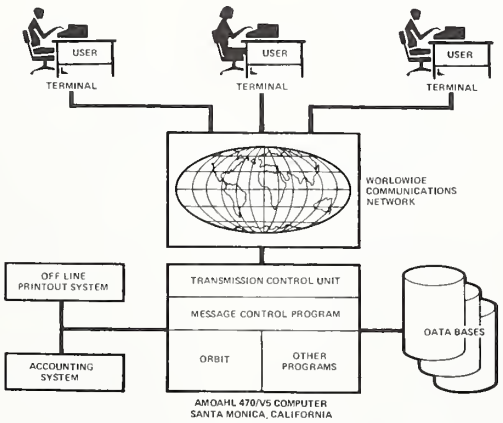


Figure 3. Schematic Representation on On-line Retrieval Service

What is the user actually doing and saying at the terminal? Figure 4 gives an example of a fairly simple search done on SDC's ORBIT system. The user types in SACCHARIN AND CANCER. The system responds by saying that there are four records in the data base that cover both these topics. The user instructs the system to print the records and it does so.

```
SS 1 /C?---SEARCH STATEMENT 1 OR COMMAND?
USER:
SACCHARIN AND CANCER

PROG:
PSTG---SEARCH STATEMENT 1---NUMBER POSTINGS (4)

SS 2 /C?---SEARCH STATEMENT 2 OR COMMAND?
USER:
"PRINT

PROG:

-1-
AN  - BA61-067735
TI  - NONNUTRITIVE SWEETENERS AND HUMAN BLADDER CANCER
      PRELIMINARY FINDINGS
AU  - KESSLER I I
SO  - J UROL (JOURA), 115 (2). 1976 143-146.

-2-
AN  - BA60-026643
TI  - COFFEE DRINKING AND CANCER OF THE LOWER URINARY
      TRACT
AU  - SIMON D; YEN S; COLE P
SO  - J NATL CANCER INST (JNCIA), 54 (3). 1975 587-592.
```

Figure 4. Sample Search in BIOSIS Data Base

This search can help to illustrate two important characteristics of on-line retrieval systems. The first is that they are fast. This search was done in Lancaster, England, which is about 7,000 miles away from the SDC Search Service computer in Santa Monica, California. From the time the first search term was entered until the last word of the results came out, the process took only 40 seconds. If someone were to perform this search using only the manual index tools available in the typical library, the search time would be measured in hours or days. Some libraries might not be able to do such a search for a patron, or their response time might be a matter of weeks.

The second important feature of on-line retrieval systems is that their use can be very inexpensive. The cost of the BIOSIS search that was just described depends in part on which retrieval system you use. Some are faster and more responsive than others and the rates for use vary somewhat from one system to another. Also, the telecommunications charges for the transatlantic connection with our computer vary from country to country. The cost for doing this search from Lancaster was $.65. If the search had been done in London, it would have cost $.53. From almost anywhere in the United States, the cost would have been about $.45.

Even in an era of limited library budgets, $.65 or even 10 or 20 times that amount is not a very high price to pay for searching about a million bibliographic records and getting exactly the information you want printed out before you in a total time of less than a minute.

The third quality of on-line retrieval systems is the ability to obtain as much precision as the user desires, *i.e.*, to obtain exactly the amount of information that is needed, with the degree of precision that is desired, without a great deal of extraneous information. This quality is a little difficult to illustrate unless one is somewhat familiar with the particular system being illustrated. But, since you have already followed one search, a somewhat more complex illustration will not be difficult to follow.

Figure 5 shows a search for a user interested in school lunch programs as related to nutrition. For this search, we will use the AGRICOLA data base.

```
SS 1 /C?
USER:
SCHOOL AND LUNCH:

PROG:
MM (LUNCH:)  (9)
          1  LUNCH (IT)
          2  LUNCH-ROOM (IT)
          3  LUNCHEON (IT)
          4  LUNCHEONETTE (IT)
          5  LUNCHEONS (IT)
          6  LUNCHES (IT)
          7  LUNCHROOM (IT)
          8  LUNCHROOMS (IT)
          9  LUNCHTIME (IT)
SPECIFY NUMBERS, ALL, OR NONE

USER:
1, 6, 9

PROG:
SS 1 PSTG (395)

SS 2 /C?
USER:
1 AND ALL NUTRITION:AND FROM 1976 THRU 1977

PROG:
SS 2 PSTG (8)
```

Figure 5. Illustration of Search-refinement Process

This user isn't quite sure how to express the "lunch" search term, so he lets the system help him. The colon after LUNCH tells the system that he is interested in documents indexed with any words beginning with LUNCH. The computer tells him in a few seconds that there are nine terms beginning that way. Which ones does he mean? The user replies that he means one, six, and nine, and the program then confines its search to the requested terms.

The user now adds the next concept--NUTRITION--to the search, and he also tells the system that he wants only documents published in 1976 or 1977. Given this more precise statement of the user's needs, the system tells him that there are three documents that meet his specification. The user can print some or all of

them, in whole or in part, or issue further expansions or refinements of his search prescription.

I hope it is clear that it is not the on-line system itself that produces precision. Rather, the precision comes from the ability of the user and system to converse with each other, letting the user inspect initial results, reformulate or refine his query, and inspect the new results. One can, of course, achieve the same kind of precision in a library or in a batch-type computer system. However, it would take much much longer and the user would undoubtedly find it embarrassing, expensive, and perhaps even dangerous to keep going back to the librarian or the batch-system operator to announce the latest revisions to his search strategy.

The fourth important feature of on-line retrieval services is that they are democratic. While many libraries have very sizeable holdings, many cities and towns throughout the U.S. and the world do not. Most federal libraries have both limited holdings and limited staff. The availability of on-line retrieval services has the potential of serving as a great equalizer between information "haves" and information "have-nots."

Some countries have been slower than others in exploring the potential of on-line information services, but the use of on-line techniques is now increasing at a very rapid rate throughout the world and most particularly in Western Europe. There are now thousands of on-line users in the United Kingdom, the Scandinavian countries, other European countries, the Middle East, Africa, Australia, Japan, South America, and on the North American continent. Most of the user organizations are libraries or information centers that serve, in total, a population of millions of scientists, technicians, and administrators. By next year at this time, on-line systems will have served millions of other users as well.

The Growth of Commercial Services

Today there are at least 15 on-line retrieval services of international scope. Of these, about 12 are based in the United States where on-line technology developed most rapidly beginning in the late 1950's.

The key technologies involved in on-line systems are time-shared computers, high-speed disk storage, and telecommunications. The computers perform the retrieval work for many users simultaneously with the cost of the computer time shared among the many users. The high-speed disk files are required to store the billions of words and numbers that make up a family of bibliographic data bases. For example, the BIOSIS Previews files stored in SDC's computer take up a total of 1,200,000,000 characters. This is the equivalent of about 200,000,000 English language words. Almost any one of those words can be located by the computer in a disk in about 20 milliseconds (1/50 of a second).

In the U.S., the early on-line retrieval systems were developed by organizations in the private or commercial sector, sometimes with government funding contributions, either in the initial development or in some phase of subsequent improvements. Most of the major on-line systems based in the United States are commercially operated. In contrast, most of the on-line retrieval services that are operating or are about to begin operating in Europe are government sponsored, in whole or in part. This difference is a function of the American tradition of free enterprise which, in the field of on-line retrieval services, has proved to be very beneficial to users.

Because the commercial on-line retrieval services based in the United States are not subsidized and do not enjoy a protected monopolistic position, they must be fully responsive to the needs of their users and offer data bases, system features, and service conditions that users want and need. Whenever one system offers a new feature or service condition that is attractive to users, the other systems must provide comparable features or services or risk losing some of their clientele to their competitors. The result of such competition has been the rapid growth of a number of important features and services.

For example, where only a few years ago some of the major on-line services were available only a few hours a day, the major U.S. systems now operate a minimum of 15 hours a day. SDC has introduced a number of other important features and services including:

Same day mailing of off-line print-outs;

The ability to truncate search terms on the right-hand side, the left-hand side, or the middle--or any combination of these--simultaneously;

On-line ordering of full-text document copies.

Document ordering capability, which we introduced in mid-1973, is now available for five of our data bases as well as for all 5,000 journal titles covered by the Institute for Scientific Information. We can fully expect a rapid expansion of this type of service now that the new Copyright Law has helped to define more clearly the rights of libraries and publishers with respect to document copying and now that there is movement towards both a periodicals clearing house and a royalties collection clearing house.

At the same time that the new system features and services have been introduced by the commercial on-line services, the range of the data bases available to the user from a single terminal has been increasing. Five years ago, the typical commercial on-line retrieval service offered users access to two or three data bases. By 1975, the number of data bases offered had grown to 15 or 20. Now, the major commercial services offer users 30 to 50 data bases and, by subscribing to several major services, users can have access to nearly 100 data bases. The users' problem is no longer one of obtaining access; it is one of selecting the combination of systems and data bases that is most appropriate for their needs or those of their clientele.[2]

One of the most important areas of change for users has been the declining cost of on-line searching. Operating a major commercial on-line retrieval service requires millions of dollars per year with much of this money going to pay for computer, disk storage, and telecommunications equipment and services. But the costs of these components are declining every year and, in spite of the fact that the data bases themselves are constantly growing, requiring more disk storage space and more computer processing time, the average cost per search for using commercial systems has held steady or actually dropped.

But the thing that keeps making commercial on-line systems better and better is not merely computers or disk drives or telephone equipment. It is the people --people in the on-line business who listen carefully to their on-line users to learn their needs and problems, people who translate these user needs into ideas for new program functions, and people who translate these ideas for new functions into the computer program code that provides the new capabilities.

It is important to note that what users wanted in 1977 was not the same as what they wanted in 1976, 1975, or 1970. As they used sophisticated new capabilities, they began to think of even more powerful ones that would make searching easier and/or more effective. So the job of on-line service organizations like SDC is not simply to listen carefully and then build a single set of capabilities. The cycle of feedback and system redevelopment is continuous and endless. That is one of the factors that make being involved in the on-line retrieval services industry fascinating and often totally exhausting.

Sophistication and power are attributes that can apply almost as much to data bases as they do to systems. Every supplier of on-line retrieval services starts from basically the same point: a magnetic tape received from an abstracting and indexing service. But, from that point on, there are profound differences in the way each on-line service designs the data base for searching and display.

This point can be illustrated with SDC's BIOSIS file. Figure 6 shows one record retrieved from our earlier search on SACCHARIN AND CANCER.

```
SS 2 /C?---SEARCH STATEMENT 2 OR COMMAND?
USER:
"PRINT 1 FULL"

PROG:

-1-
AN  - BA61-067735
TI  - NONNUTRITIVE SWEETENERS AND HUMAN BLADDER CANCER
      PRELIMINARY FINDINGS
AU  - KESSLER I I
SO  - J URON (JOURA), 115 (2). 1976 143-146.
CC  - #24007 (NEOPLSMS/NEOPL AGNTS-CARCINOGENS); 07004
      (BEHAVIOR BIOL-HUMAN BEHAVIOR); 10060 (BIOCHEM
      STUD-GENERAL); 10068 (BIOCHEM STUD-CARBOHY-
      DRATES); 12503 (PATHOLOGY-COMPARATIVE); 12504
      (PATHOLOGY-DIAGNOSTIC); 13004 (METABOLISM-
      CARBOHYDRATES); 13020 (METABOLISM-METABOLIC
      DISORDERS); *13532 (FOOD TECH-PREP, PROCESSING,
      STORAGE); 15501 (URIN SYST/EXTERN SECR-GEN STUD,
      METH); *15506 (URIN SYST/EXTERN SECR-PATHOLOGY);
      17008 (ENDOCRINE SYST-PANCREAS); 19001 (CENTAL/
      ORAL BIOL-GENL STUD, METHS); 21004 (PSYCHIATRY-
      ADDICTION (INC SMOKING)); 22100 (ROUTES OF
      IMMUNIZ,INFECT, THERAP); 22501 (TOXICOL-GENL/
      EXP STUDS,METHS); *22502 (TOXICOL-FOOD,RESIDS,
      ADDIT,PRESRV); 24001 (NEOPLSMS/NEOPL AGNTS-
      DIAGNS METH); 51522 (PLANT PHYSIOL-CHEM
      CONSTITUENTS)
TC  - 11000 (PLANTAE-UNSPECIFIED); 86215 (HOMINIDAE)
IT  - SACCHARIN CYCLAMATE CARCINOGENS SMOKING
```

Figure 6. Sample BIOSIS Search (Full Citation)

You will recall that we found four relevant citations. If we examine this citation carefully, we notice that there are a number of other terms that might be, and perhaps should have been, used as search terms for our SACCHARIN AND CANCER search. For example, the title reminds us to consider the concept of sweeteners or sweetening; the index-term field makes us consider the idea of expanding our search to include cyclamates and to use the word carcinogens in addition to cancer; and a field we call Category Codes, based on the BIOSIS Taxonomic Codes, tells us that using the code 24007 will help us to locate literature on neoplasms, neoplastic agents, and carcinogens. These words do not appear on the standard tape that we receive from BIOSIS. We deliberately create this field, using dictionary information provided to us by BIOSIS on a totally separate tape. Doing so expands our cost of processing and storage but it provides an invaluable aid to the user, who can learn directly and immediately the meaning of the Category Codes he sees and can translate this information into a more successful search.

Figure 7 shows a summary of the new search.

```
SS 6 /C?
USER:
"HISTORY

PROG:
SS  1:  SACCHARIN AND CANCER  (4)
SS  2:  SACCHARIN OR ALL SWEETEN: OR ALL CYCLAMATE:
        (326)
SS  3:  CANCER OR ALL CARCINOGEN: OR (33869)
SS  4:  2 AND 3  (38)
```

Figure 7. Summary of BIOSIS Search

Search Statement 1 was our original search which netted four citations. New Search Statement 2 expands the SACCHARIN concept to include all terms beginning with SWEETEN or all terms beginning with CYCLAMATE. This gives us 326 documents. In Search Statement 3, we expand the CANCER concept to include all terms beginning with CARCINOGEN and we also include the category 24007 which we discovered a moment ago. The result of this Search Statement is 33,869 citations. The final result is 38 citations, over nine times as many as our initial simple search. I am sure that a searcher familiar with biological literature and nomenclature would be able to locate even more relevant citations with no difficulty.

The Future

What can we expect to see in the future? In the near future, the answer is "more of the same." There will be more data bases, more systems, longer service hours, faster print-out service, lower prices per search, and additional system features. There will also be more new demands and challenges for the searcher.

For example, the literature of on-line searching indicates that, because of the conceptual overlap among data bases, very few searches can be done on only one data base. More often, two, three, or more data bases must be searched to satisfy the needs of the client. As the number of on-line data bases grows and as clients become more demanding about exhaustive

searches, the task of the searcher will become increasingly difficult.

One of the family of tools that are being developed to help the searcher meet such new challenges is the capability to take a basic search strategy that is developed in one data base and execute it in one or more other data bases without having to re-key the terms. Figures 8, 9, 10, and 11 illustrate the process.

```
SS 1 /C?
USER:
"SAVE LUNCH

PROG:
ENTER SEARCH-

*SVS* SS 1 /C?
USER:
LUNCH OR LUNCHES OR LUNCHTIME

PROG:
SS 1 PSTG (555)

*SVS* SS 2 /C?
USER:
1 AND ALL SCHOOL: AND ALL NUTRITION:

PROG:
SS 2 PSTG (92)

*SVS* SS 3 /C?
USER:
2 AND ALL PROGRAM#

PROG:
SS 3 PSTG (21)

*SVS* SS 4 /C?
USER:
"PRINT TI

PROG:

-1-
TI  - EVALUATION OF SCHOOL LUNCH AND SCHOOL BREAKFAST
      PROGRAMS IN THE STATE OF WASHINGTON

-2-
TI  - REPORT ON THE SCHOOL LUNCH AND BREAKFAST PRO-
      GRAMS IN WASHINGTON, D.C.

-3-
TI  - EVALUATION OF SCHOOL LUNCH AND SCHOOL BREAKFAST
      PROGRAMS IN THE STATE OF WASHINGTON

-4-
TI  - CHRONOLOGICAL LEGISLATIVE HISTORY OF CHILD
      NUTRITION PROGRAMS/ /UNITED STATES DEPARTMENT
      OF AGRICULTURE, FOOD AND NUTRITION SERVICE,
      BUDGET AND PLANNING DIVISION. --

-5-
TI  - A COORDINATED NUTRITION EDUCATION ACTION PROGRAM
      IN THE ELEMENTARY CLASSROOM, THE SCHOOL LUNCH
      ROOM, AND THE HOME
```

Figure 8. Establishing a SAVEsearch on the AGRICOLA Data Base

```
SS 4 /C?
USER:
"FILE CIS

PROG:
ELAPSED TIME ON AGRICOLA:  0.10 HRS.
YOU ARE NOW CONNECTED TO THE CIS INDEX DATABASE.
COPYRIGHT CIS, INC., 1977

SS 1 /C?
USER:
"RECALL  LUNCH""PRINT INDENTED"

PROG:
SS  1:  LUNCH OR LUNCHES OR LUNCHTIME  (158)
SS  2:  1 AND ALL SCHOOL: AND ALL NUTRITION:  (117)
SS  3:  2 AND ALL PROGRAM#  (114)

-1-
ACCESSION NUMBER          77-S163-9
TITLE(MAIN)               CHILD NUTRITION ACT OF 1977.
SOURCE                    JUNE 16, 1977.; 94 P.
AVAILABILITY              ^ CIS/MF/3
ISSUE(UPDATE CODE)        7706
CONGRESS & SESSION NO.    95-1
SUPT. OF DOC./ITEM NOS.   ITEM 1008-A
DOC/RPI/BILL NOS.         S. RPT. 95-277

-2-
ACCESSION NUMBER          77-H343-10
TITLE(MAIN)               NATIONAL SCHOOL LUNCH ACT AND
                          CHILD NUTRITION AMENDMENTS OF
                          1977.
SOURCE                    MAY 10, 1977.; 84 P.
AVAILABILITY              ^ CIS/MF/3
ISSUE(UPDATE CODE)        7705
CONGRESS & SESSION NO.    95-1
SUPT. OF DOC./ITEM NOS.   ITEM 1008-A
DOC/RPT/BILL NOS.         H. RPT. 95-281

-3-
ACCESSION NUMBER          77-H341-23
TITLE(MAIN)               OVERSIGHT HEARINGS ON THE
                          SCHOOL LUNCH PROGRAM.
SOURCE                    FEB. 20, JUNE 17, 24, 30,
                          JULY 22, 24, AUG. 2, 9, 25, 26,
                          30, 31, SEPT. 1, 2, 30; VIII+
                          784 P. IL.
AVAILABILITY              ^ CIS/MF/10
ISSUE(UPDATE CODE)        7704
CONGRESS & SESSION NO.    94-2
SUPT. OF DOC./ITEM NOS.   Y4.ED8/1:L97/4. ITEM 1015


MONTHLY CAT./LC NOS.      77-601140.
*-*-*
ACCESSION NUMBER          77-H341-23.1
ANALYTIC RECORD SIZE        816
ANALYTIC SOURCE           FEB. 20, 1976. P. 4-48
FIRST WITNESS             BEVINS, C. E. (AND OTHERS)
```

Figure 9. SAVEsearch: CIS Index

```
"FILE CRECORD""RECALL LUNCH"

PROG:
ELAPSED TIME ON CIS INDEX:  0.07 HRS.
YOU ARE NOW CONNECTED TO THE CRECORD DATABASE.
```

```
NP (LUNCHTIME)
SS  1:  LUNCH OR LUNCHES OR LUNCHTIME  ((*)
SS  2:  1 AND ALL SCHOOL: AND ALL NUTRITION:  (48)
SS  3:  2 AND ALL PROGRAM#  (33)

SS 4 /C?
USER:
"PRT INDENTED

PROG:

-1-
ACCESSION NUMBER        7732-1147
TITLE(ABSTRACT)         CONFEREES TO MEET ON HR 1139 TO
                        REVISE AND EXTEND THRU FY80
                        CHILD NUTRITION PROGRAMS ADMINIS-
                        TERED UNDER THE NATL SCHOOL LUNCH
                        ACT AND CHILD NUTRITION ACT; SEPT
                        27, 1 PM, S-126 CAPITOL.
REFERENCE               SEPTEMBER 23, 1977P. D1347
CONGRESS & SESSION      95-1 (149)
CATEGORY CODES          4310 (SOCIAL SERVICES: FOOD &
                        NUTRITION)

-2-
ACCESSION NUMBER        7732-581
TITLE(ABSTRACT)         HOLLINGS SUPPORTS CONTINUATION OF
                        C/NUTRITION, DISCUSSES SPECIFIC
                        PROGRAMS OF THE COMMITTEE, IN-
                        CLUDING WIC, SCHOOL LUNCH, CHILD
                        CARE, COMMODITY SUPPLEMENTALS,
                        AND SUMMER FOOD; CITES CRS STUDY
                        WHICH STATES THAT COMPREHENSIVE
                        INFORMATION FOR DETERMINING
                        DIRECTION OF FEDERAL HUMAN NUTRI-
                        TION RESEARCH IS LACKING.
REFERENCE               SEPTEMBER 21, 1977P. S15308.
CONGRESS & SESSION      95-1 (147)
CATEGORY CODES          3012 (HEALTH: RESEARCH);  4310
                        (SOCIAL SERVICES: FOOD & NUTRI-
                        TION)
```

Figure 10. SAVEsearch: CRECORD

```
-1-
ACCESSION NUMBER            GY  70763
TITLE                       SYNTHESIS OF CRITERIA FOR
                            THE DEVELOPMENT OF CHILD
                            NUTRITION PROGRAMS
INVESTIGATORS               HARD MM; PURCE DW
ORGANIZATIONAL SOURCE       WASHINGTON STATE UNIVERSITY,
                            SCHOOL OF HOME ECONOMICS,
                            HOME ECONOMICS RESEARCH,
                            PULLMAN, WASHINGTON, 99163
PERIOD OF PERFORMANCE       10/76 TO 9/77
FISCAL YEAP                 77
SPONSORING ORG.             U.S. DEPT. OF AGRICULTURE,
                            COOPERATIVE STATE RES.
                            SERVICE, WASHINGTON
SPONS. ORG. CONTROL NO.     0070763; WNP00331
TECHNICAL SUMMARY (AB)

OBJECTIVE:  INTERRELATE THE TYPES, AMOUNTS AND NUTRI-
TIONAL CONTRIBUTIONS OF FOOD CONSUMED, FOOD PREFER-
ENCES, HOUSEHOLD FOOD PATTERNS, ETHNICITY, PHYSICAL
CHARACTERISTICS, SOCIOECONOMIC FACTORS, BIOCHEMICAL
PARAMETERS OF 8-12 YEAR-OLD CHILDREN IN SCHOOL LUNCH
PROGRAMS, DELINEATE STANDARDS FOR ASSESSING NUTRI-
TIONAL STATUS OF THESE CHILDREN.
```

Figure 11. SAVEsearch: SSIE

Assume that we are already connected to the AGRICOLA data base. With the SAVE Command, we have asked ORBIT to save all of our subsequently entered searches, under the name LUNCH. ORBIT places us in the SAVE-search mode and we continue to enter each of the search statements:

> In SS1, we enter those terms, LUNCH, LUNCHES, AND LUNCHTIME, in an ORed statement, to retrieve any citations that contain one or more of the terms. You may remember from Figure 5 that we could, as well, enter LUNCH: and select our search terms from the display of terms that start with LUNCH as a stem;
>
> In SS 2, we request that those 555 citations be matched against citations that contain any term beginning with SCHOOL, and with any terms beginning with NUTRITION. The subset of our initial 555 citations meeting those specifications is 92;
>
> In SS 3, we narrow down the results to those citations that contain either PROGRAM or PROGRAMS. The pound symbol stands for one character--either a blank or a letter--and, in effect, permits us to request both singular and plural forms of any term. The ALL before each of the truncated entries in this series tells ORBIT, in advance, that we want all of the terms beginning with the entered stem to be ORed together, without our first seeing the display of terms.

Before settling on this search strategy, we review some of the titles by issuing a PRINT TI command. If we are satisfied that the strategy is adequate, we tell ORBIT (not shown) that we have finished entering this SAVEsearch.

At this point we have several options. We can print some or all of the citations at the terminal, or have them printed off-line, for mailing to us that day. We will skip these print-outs and continue with our illustration of the search-saving technique. This topic of nutrition and school lunch programs has been of concern in Congress, so we could expect to find a wealth of information in the Hearings, Committee Prints, and Reports that have emanated from the congressional testimony and investigatory activities.

Because we have saved our strategy under the name LUNCH, we can now ask to be connected to the CIS (Congressional Information Service) data base and then RECALL--that is, apply--our previous search strategy. Figure 9 shows how ORBIT re-executes the search. It also shows the number of CIS citations for each of our searches. At this point, the searcher might review some of the citations to identify controlled vocabulary terms that are particularly apt for the search, *e.g.*, the name of a particular Bill or Act, and revise the strategy accordingly. However, for our purposes here today, we are simply going to show, in an abbreviated format, three of the citations that resulted from this search, so that you can see the various different kinds of bibliographic information that are associated with this data base: the Congress

and Session number; the Superintendent of Documents and Item numbers; any relevant Report numbers. There is also available a detailed abstract which could be displayed by saying "PRINT FULL."

The first two documents in Figure 9 are Reports. The third document is a Hearing. In the Hearing record, we are given a reference to a specific group of witnesses who discussed the nutritional aspect of the school lunch program. The sub-record, with;its own page-range references, is shown after the asterisked-dashed line, and is a specific part--77-H341-23.1--of the main record, OVERSIGHT HEARINGS ON THE SCHOOL LUNCH PROGRAM.

We might want to pursue our interest in Congressionally-related information on this topic by searching still another data base--the Congressional Record Abstracts. This data base, which is based on the *Congressional Record*, tells us what has happened on this topic on the floor of Congress. We follow the same procedures, this time requesting the CRECORD data base. We are told that there were no citations posted --that is, NP--to the term LUNCHTIME, but the other terms in our search, LUNCH OR LUNCHES, yielded 98 citations. The remainder of our strategy is given and we learn that there are 33 citations.

In reviewing the first few records from this search, we find the dates and particular page numbers of the complete discussions in the *Congressional Record*. We also learn that there is a code, 4310, that represents the concept of SOCIAL SERVICES: FOOD & NUTRITION. This code may be a useful supplementary search term for our strategy and the searcher wanting to develop a comprehensive search strategy could include this term, on the spot, in a revised approach.

Our final example, in Figure 11, shows one result of applying our SAVEsearch to the SSIE (Smithsonian Science Information Exchange) data base, which covers current research. We are told the investigators, the organization doing the research, the period of performance of this research, and the sponsoring organization. We are also given a summary of the research in progress.

The search of four data bases does not exhaust the number of possibilities in tracking this particular search topic but it should serve to illustrate the power and usefulness of search-saving techniques and the help that such techniques can give to searchers in coping with a multiplicity of information sources.

Looking further ahead, we can expect to see a leveling out of capabilities, from one system to the next, as the higher speeds of computers erode some of the advantages in response time now enjoyed by the better designed and more efficient retrieval programs. When this happens, more and more emphasis will be placed on the quality of customer service and on the things that make some systems easier or cheaper to use than others, or that provide added dividends to the user.

One highly important type of system improvement will come increasingly into view in the next several years --the personalized system. The personalized system will let the user establish a "profile" of the special way in which the system is to behave for that user.

The computer will remember how each user wants the system to behave and how to converse and will automatically adopt the mode that the user has requested, every time the user logs in. Language is one way in which a system can be personalized by the user for his or her use. Another is screen size. Video-display terminals are becoming less and less expensive and more and more portable, but the more portable terminals generally have a smaller screen capacity. The user may need to ask the system to output lines of 40 characters width or 47 or 63, or any number, instead of the standard length of 72 characters usually used for standard printer-type terminals.

Of course, there will be many other changes that will simplify the use of on-line systems. We will certainly have on-line aids to help the searcher select data bases from the hundreds that are available. There will almost certainly be mechanisms for translating the user's entry terms into controlled vocabulary terms in the data base. And we will probably see mechanisms for translating the terms in one system's command language into the terms required by another system with which the user is less familiar.

We can also expect to see greater vertical integration within the information retrieval services industry, with a blurring of the line that exists between data base publishers and on-line service suppliers. The present relationship between these two groups is very complex and will become more so before a simpler and more stable long-term relationship begins to emerge.

Conclusion

Many of the present users of on-line services have been witnesses to the birth and growing pains of these services. After all, such services have grown up only within the past 12 years. Because on-line technology is growing so rapidly, many of today's users will also witness most if not all of the new tools, mechanisms, and trends we have been reviewing. Some of these will be here within a matter of months. Others will take years to reach operational status. Whatever the rate of maturation, on-line retrieval services are probably the most fascinating and pervasive development in the field of library and information science in the past decade. They have significantly raised our informational standard of living.

REFERENCES

1. Williams, Martha E. "Education and Training for Online Use of Data Bases," presented at EUSIDIC Conference, Graz, Austria, December 1, 1976.

2. Wanger, Judith, Cuadra, Carlos A., and Fishburn, Mary. *Impact of On-line Retrieval Services: A Survey of Users, 1974-5.* Santa Monica, Calif.: Systems Development Corporation, 1976.

EFFECTIVE TRAINING: THE KEY TO EFFICIENT RETRIEVAL

by

Charles L. Gilreath*

The decades of the sixties and the seventies have been a period of intensive activity in application of computer technology to library operations. Nowhere has the activity been more dramatic than in the development of on-line computer systems for both the creation and the searching of bibliographic records. The sixties and seventies have given us, for instance, a whole new vocabulary of acronyms--OCLC, MARC, BALLOTS, CAIN, and AGRICOLA, to name a few. Martha Williams has recently estimated that there are at least 277 machine readable bibliographic data bases available in the world. These computer files contain over 52 million individual records.[1] If these figures are anywhere close to accurate, the National Agricultural Library's Agricultural On-line Access data base, AGRICOLA, represents about two percent of the total number of machine processed records available today and it is the largest single source of computerized bibliographic records in the fields of agricultural research.

The advent of computer searching programs for retrieval of published literature has had a significant impact on the services provided in libraries and information centers. In organizations such as special libraries, computer searching has enabled staff members to increase their productivity greatly by providing them a fast and comprehensive means of searching relevant literature for their clients in research and development. In other organizations--notably academic libraries--the availability of computerized data bases has made possible the addition of both retrospective and current awareness literature searching to individual faculty members and students, a level of service not previously possible on any but the most limited scale, if at all.

The success of batch mode current awareness programs such as CALS, the Current Awareness Literature Service provided to USDA scientists through the USDA Agricultural Research Service, and the phenomenal growth of firms such as Lockheed Retrieval Service and Systems Development Corporation are clear evidence of the dramatic changes which have come to libraries. The improvements in information services produced by data bases have, however, created a whole new set of problems with which we must wrestle. Many of the issues that have been raised in the literature regarding searching of data bases have centered on the costs, administration, and efficient use of such services in libraries.

No bibliographic tool in recent times has produced as much cost consciousness among librarians as have these services which are billed on the basis of actual use. For example, the subscription rates for individual printed research tools such as *Science Citation Index* or *Chemical Abstracts* may be frightfully high and the overall cost for maintaining even the most basic reference collection will certainly run into several tens of thousands of dollars each year. They are not, however, matters which daily occupy the minds of the librarians and information specialists who are providing service out of their pages. Concerns over the costs of traditional bibliographic tools are usually relegated to that most pitiable of creatures, the library administrator. He may in his misery, of course, raise the issue of cost from time to time but it is not a frequent topic in discussions relating to the daily provision of reference service. Computer search systems, on the other hand, bring the issue of costs home on a regular basis not only to the administrator who must pay the monthly bills for usage but also to the search analyst who receives cost related data at the end of each terminal session and to the library patron who not uncommonly will receive an invoice for at least part of the costs of a search. Such heightened awareness of the price tag on information is probably a healthy thing both for the information specialist and for his client. Unlike printed bibliographic tools where monitoring of the per user costs are extremely difficult, computerized data base services present the information specialist a mass of hard data with which to calculate rational cost-benefit analyses of his services. And--without getting into the whole issue of whether passing charges on to users is philosophically, morally, ethically, or fiscally right or wrong--it can also be conceded by most that it is probably a good thing that the client can see just how much information does cost.

The second major issue that has surfaced over the past few years is the matter of just who will provide the literature searching service and what will be the nature of the service. Will highly qualified subject specialists interact with the computer to produce an initial list of references which are then edited to produce a more refined list of highly relevant citations for the patron or will the general reference staff act as the interface to produce the initial list which the patron screens himself? Both modes of operations have been reported in the literature but the most common method appears to be the second alternative--the general reference staff member or search analyst is usually a well trained professional who can ask the right questions but who may not have the expertise to edit every print-out that he produces. These two issues--cost and the intellectual end of information retrieval--bear directly upon the central topic of this paper, that is, the role of training and education in the use of computerized literature searching services.

While one should never slight the importance of good training in any area of reference work, whether the tools being used are computer data bases or printed indexes and abstracting services, the issue is particularly important for the efficient retrieval of information by machine. Assuredly, there is a definite cost involved when a librarian or information specialist fails to negotiate any reference query properly, but failure to approach the right bibliographic tool for the question or slowness in working through a printed index does not generally result in the patron's being handed an invoice for $10 or more, as is the case with computer search services which pass their

*Charles L. Gilreath, Head, Automated Information Retrieval Service, Texas A & M University Library.

costs on to the user. Because his actions do result in a direct cost to his patron, the search analyst feels more acutely the need for good training in the skills appropriate to efficient retrieval of information from machine readable files.

Training for computer retrieval of literature covers a great expanse of territory. Questions as wide-ranging as who should be the object of the training programs to who should be providing such training or what techniques are most effective for teaching terminal operation are all legitimate and important issues for discussion, yet one cannot hope to raise all relevant issues on this topic, let alone attempt an answer to them. This paper will, therefore, focus on three major questions related to the training of those who will function as search analysts. Those questions are: (1) What skills are necessary for efficient retrieval of bibliographic data by computer? (2) What are the roles of various organizations in providing on-line training? and (3) What are the elements of a good training program?

Nature of Skills Required

The skills required for success in the use of on-line bibliographic retrieval systems can be subsumed under three broad headings: interviewing skills, search strategy formulation skills, and system operation skills. These skills are summarized in the *Cain Online User's Guide* this way:

> Efficient retrieval from a computerized bibliographic data base requires 1) a clear idea of the topic, clearly stated; 2) the development of an adequate list of key words divided logically into concept groups; 3) an ability to adjust the initial strategy on the basis of information gained from preliminary review of search output.[2]

Although each aspect will be discussed separately here, they should be considered as interconnected parts of a whole. The interview process is an integral part of the search strategy; when performed correctly it is, in itself, an art. To some degree this is the essence of a good reference librarian or search analyst--that is, a person who can elicit from another the proper description of what information is desired. This ability to ask the right questions cannot, properly speaking, be taught in a training program. But the nature of the questions which must be answered in order to formulate a successful computer search strategy can be taught. The interview with the patron of a computer search service is, of course, just another form of a good reference interview: the information specialist is seeking a clear idea of just what kind of information is desired, what aspects of the topic are of no interest, and some idea of how much information is required to answer the query. The computer search interview differs from the normal reference interview in its emphasis on generating a list of words or other data elements which, in conjunctions with each other, will produce a list of relevant documents. Because most computerized bibliographic data bases rely not only on the controlled vocabulary applied by indexers and catalogers but also on terms occurring in titles and abstracts, success in retrieval from these systems relies more heavily on the user's knowledge of variant word forms and synonymous terms. While it may be adequate, for example, to approach a good printed index with the term "cattle" and rely on the consistency of the indexer or the syndetic structure of the bibliographic tool to lead to related terms, a similar approach to a computerized file would be ill-advised. No fewer than 15 different terms would be required to retrieve the relevant literature on this topic in a data base such as AGRICOLA.

Formulation of an adequate search strategy to retrieve information on the topic described in the interview process is the next step in the search procedure. Three types of knowledge must be acquired in order for the search analyst to be successful in this aspect of his task: 1) knowlege of data base content and structure; 2) familiarity with principles of information retrieval via computer; and 3) awareness of protocols of the retrieval system as they relate to specific aspects of the data base. These topics plus actual operation of specific retrieval systems are the bulk of the various training programs now being conducted.

Early training programs, as Janet Egeland has pointed out, tended to emphasize principles of retrieval and system operation, giving only passing reference to data base content and structure.[3] As users got over their initial awkwardness with using the computer systems, however, they began to feel more acutely the need for better information on the way in which the files they were trying to search were put together. The vendor manuals, which contained brief summaries on the various data bases and concentrated on protocols for retrieving information from various fields of the unit record, were simply inadequate to explain policies followed by data base producers for inputting records into their files.

This need for information about input policies and procedures created something of a crisis among data base producers, a crisis that has not wholly abated even today. In order to provide the kind of information desired by on-line users, many data base producers had to pull together facts about their policies and procedures which perhaps existed only in scattered memoranda and meeting minutes or which had been handed down, epic-like, in an oral tradition. The National Library of Medicine has perhaps led the field in developing materials to acquaint users with their machine readable files and the content and structure of their data bases. Others such as the National Agricultural Library, Biosciences Information Service (BIOSIS), Psychological Abstracts have quickly followed suit in producing detailed guides to their files. Other data base producers such as the National Technical Information Service and the Commonwealth Agricultural Bureaux have been slower about developing adequate guides but the message now appears clear that input and retrieval are but the two sides of the information coin; while it is not necessary for a reference librarian or search analyst to be trained as an indexer, it is highly desirable that the individual know a great deal about the conditions under which records are created for the files that are being searched.

Training users in the principles of on-line retrieval involves instruction in coordinate indexing, in Boolean logic, and in a generalized model of a computerized information retrieval system. The instruction in the use of the Boolean logical operators OR, AND, and NOT is clearly a necessity for any

retrieval program for, without an understanding of these operators, the user is doomed to frustration and failure. The importance of the other two elements--coordinate indexing and a generalized computer retrieval model--is perhaps overlooked more than it should be. To be sure, the main goal of most training programs (as opposed to a more comprehensive educational program in information retrieval) is not to make expert computer programmers or systems analysts; yet without a clear understanding of what a computerized bibliographic program is actually doing--albeit at a highly generalized level--the user will frequently be handicapped in his ability to analyze his mistakes or to understand certain responses given him by the computer.

Finally, search strategy formulation requires that the search analyst understand the various protocols for entering search parameters in a given access program. This knowledge differs from pure access language features and refers to the analyst's understanding of how one vendor's version of a bibliographic file differs from that of another. A data base producer such as the National Agricultural Library creates a standardized product which is delivered to those who have purchased a copy of the sale tapes. In the process of reformatting these standard records for use on their own search system, vendors create a new record. For example, in the AGRICOLA unit record there is enough space for entry of two subject category code numbers; one on-line vendor merges these subject codes into one searchable field while the other has chosen to put these codes into two separate fields. The search analyst who is not aware of these differences between the two systems may on occasion fail to retrieve a sizable portion of the literature in the data base that is relevant to his search topic.

Having elicited a clear statement of the patron's search request during the interview process and having formulated what appears to be an efficient strategy to retrieve information on the topic, the search analyst must employ another set of skills in actually operating the equipment and using the retrieval program interactively. This aspect of training is the least amenable to pure intellectualizing; facility in using the system is clearly a function of the amount of hands on time a user has invested. Over the past two years, the National Agricultural Library has sponsored six week-long training sessions on use of the AGRICOLA data base. Among participants in these workshops, roughly 30 participants were novices at computer searching. It has been my observation that among these people new to the whole business of data base manipulation, their greatest sense of frustration came from actual manipulation of the retrieval systems. Review of proficiency exercises administered at the end of each workshop clearly indicates that most participants knew reasonably well the content and structure of the file and could formulate on paper a logical strategy to retrieve requested information. Actually getting the written strategy converted to results and then adjusting the initial strategy as necessary while in an interactive mode is, without doubt, the hardest part of the whole process for new searchers.

While an on-line session may appear to the casual observer as simple as typing a letter, it is the culmination of many separate operations and it demands almost simultaneous use of many skills. The search analyst will be working out on the terminal the details of a strategy that may have taken an hour or more to develop. In the process, he is having to keep track of individual concepts in the form of set or search statement numbers, combine these concepts with Boolean logical operators, review titles for relevance to the topic, and add or delete terms in order to adjust the results of the search. At the same time he may have the additional elements of a patron sitting next to him who wants to know what's going on and the constant hum of the terminal quietly reminding him that several million dollars worth of computer is connected to his terminal and the meter is ticking. In reflection on the skills necessary to conduct an adequate search of the literature on a data base such as AGRICOLA, I might paraphrase Dr. Johnson's rather ungracious statement about women preachers: The amazing thing is not that we can train people to search well but that we can train them to do it at all.

Organizational Roles in Training

Training in the use of bibliographic data bases is provided by a number of different organizations which can be grouped into the three broad categories: on-line vendors, data base producers, and professional/educational organizations. The categories are not that neatly defined in actual practice since some organizations such as the National Library of Medicine function as both a data base producer and an on-line vendor, but they will provide a framework for a discussion of the types of training which each group is best able to provide.

The on-line vendors have, as one might reasonably surmise, been very active in providing training programs for users and potential users. Their primary strength lies in training searchers in the skills of manipulating the access language and in familiarizing users with the basic structure of their version of a data base. Because they have been active in the development of the data base for on-line searching, they are in a particularly good position to explain how various elements of the data base are linked to each other in the on-line mode. By providing this fundamental information and by allowing trainees rather generous access to the data base for experimentation, they are able to help novices over that initial hurdle of familiarizing themselves with operating equipment and manipulating terms.

In addition to formal training programs, on-line vendors have been active in exploring other avenues for providing users with assistance. Built into each access program are numerous tutorial features which the user can call up whenever he encounters a problem that he cannot resolve quickly by looking in the vendor's manual. These features vary from simple commands such as "Help" to detailed explanations of particular system features. Most systems have gone beyond these on-demand tutorial programs and have provided other on-line training aids. SDC, for example, has developed from the records in the NTIS data base a demonstration file which a beginning searcher might use to practice on for a while. The National Library of Medicine and Lockheed have taken this concept further and have developed a series of exercises through which a novice searcher can work in order to learn about system operation. Lockheed's ONTAP provides exercises for the searcher and an assessment of their results. MEDLEARN goes beyond this and actually

presents the new MEDLINE searchers-in-training with an entire learning package.

After the initial discomfort with actual system operation has subsided and the searcher feels relatively comfortable with formulating computer strategies, his level of awareness about the nature of the data bases he is searching will become more acute. While on-line vendors may attempt to cover content and structure of the individual files which they support, they cannot provide authoritative answers to every question that may arise about a particular data base. This is the point at which the data base producer must become active in the training function. Some producers have tried to meet the perceived needs of on-line searchers by publishing manuals and other informational brochures, while others have been more aggressive in developing their own training programs. The scope and purpose of these programs varies from half-day lecture presentations in which the data base is described to the more rigorous training courses of NAL and NLM in which participants not only are exposed to lectures on the input policies and practices but also are given extensive on-line time to work through problem sets designed to illustrate data base features.

Other organizations involved in the training of search analysts range from professional associations to schools of library science and, on a more informal level, on-line user groups. Many of the activities of these groups tend to place less emphasis upon the strict interpretation of training and concentrate more on what may be defined as education for information retrieval.[4] Within these groups one encounters formal courses in librarianship which cover in detail the theory and practice of data base construction and searching as well as issues relating to the management of computer search services within the setting of a library or information center. Professional organizations frequently serve both as forums for discussion of management issues and as sponsors of many discipline-oriented on-line workshops. Because of their more formal structure and the number of persons usually involved, programs sponsored by schools and professional organizations cannot always provide training opportunities for more day-to-day situations. Partly as a result of the need for a more responsive mode of sharing information, several on-line users groups have been organized in the past two years. These groups are usually composed of search analysts within a fairly prescribed geographical area who meet periodically to share information they have gathered through their own on-line experiences and to discuss common problems. Some of the programs of these groups have been really quite imaginative. For example, one on-line users group in Texas met to test a group of new data bases offered by one vendor. Members agreed to donate the free hour's on-line time provided them by the vendor for this program, so the group members had a chance to conduct a rather extensive peer training session at minimal cost.

Elements of a Training Program

Designing a training program for on-line searching can be a time-consuming and frustrating task unless one first resolves several basic issues such as program objectives, time frame, and budget. Once those issues are resolved the job is still one of fitting pieces of instructional content into an overall sequence which will aid participant learning and balance modes of presentation so that those with various learning styles can have an opportunity to acquire the workshop content.

Although writing down specific objectives for an on-line workshop may not be high on an instructor's list of enjoyable things to do, it is perhaps the clearest method available of quantifying how much material needs to be covered in the time allotted. Actually seeing statements of objectives to be accomplished during a workshop can be helpful in dividing instructional sessions into logical units and can often suggest methods of presentation as well, especially if the objectives to be accomplished are stated in behavioral terms.

A fairly clear idea of the budget for a training session is also critical, for it will bear directly upon the modes of presentation to be used. To be sure, a minimal budget will dictate that relatively greater emphasis will have to be placed on lecture/demonstration presentations and less on hands-on exercises with participants. A small budget does not, however, mean that variety cannot be introduced into a presentation. For example, a very effective simulation of an on-line search sequence can be made using nothing more sophisticated than an overhead projector and a couple of transparencies.

Once objectives, budget, and other such basic issues are settled, development of specific parts of the presentation can begin. In this context one needs to keep in mind the point made by Lancaster in his book *Information Retrieval On-line*, that being, "The fact that no single instruction method or booklet, no single 'style' of presentation, no single compromise between brevity and completeness, seems to satisfy all, or even a majority of users."[5] A second point also needs to be remembered as well: The most effective learning comes about when the learner is active--in other words, we learn by doing. It therefore behooves the trainer to aim for a mixture of lecture and hands-on exercises that will allow the participants to check their progress periodically. Check-lists, or other pencil and paper or oral exercises are often effective methods of both breaking the monotony of a lecture and providing valuable feedback close to the point of primary stimulus.

In programs where access to the on-line system is provided trainees, a decision needs to be made by the instructor what use will be made of the time available. Will participants be asked to work through a predetermined set of problems designed to illustrate data base features or will the on-line time be spent helping participants work through questions they have thought of during the lecture or--more typically--have brought with them? Both methods can be helpful in teaching but past experience with this type program has led me to the conviction that a well-developed set of exercises which make use of salient system and data base features will, in the long run, be much more satisfying to those being trained. All too often a participant who has been told to experiment with the data base on his own will come away from the session wishing that he had seen more system features than he had been able to think of on his own. Evaluation forms from the AGRICOLA workshops held over the past two years seem to bear out this contention. Out of over 90 participants in five workshops only about three

have answered yes to the question, "Do you think you would have learned as much about the data base if you had been allowed to do problems of your own choosing?"

Conclusion

Much else could be said about training persons as searchers of computerized bibliographic data bases but that would be a task much greater than is appropriate for this brief paper. By way of a conclusion, however, I will just leave you with a word taught me by a finance professor: TINSTAAFL, there is no such thing as a free lunch. As applied to computer literature searching services it means simply this, either we spend the time and effort in training searchers in efficient retrieval techniques or most assuredly we shall pay greater costs when using the systems available to us, for the meter is indeed ticking.

REFERENCES

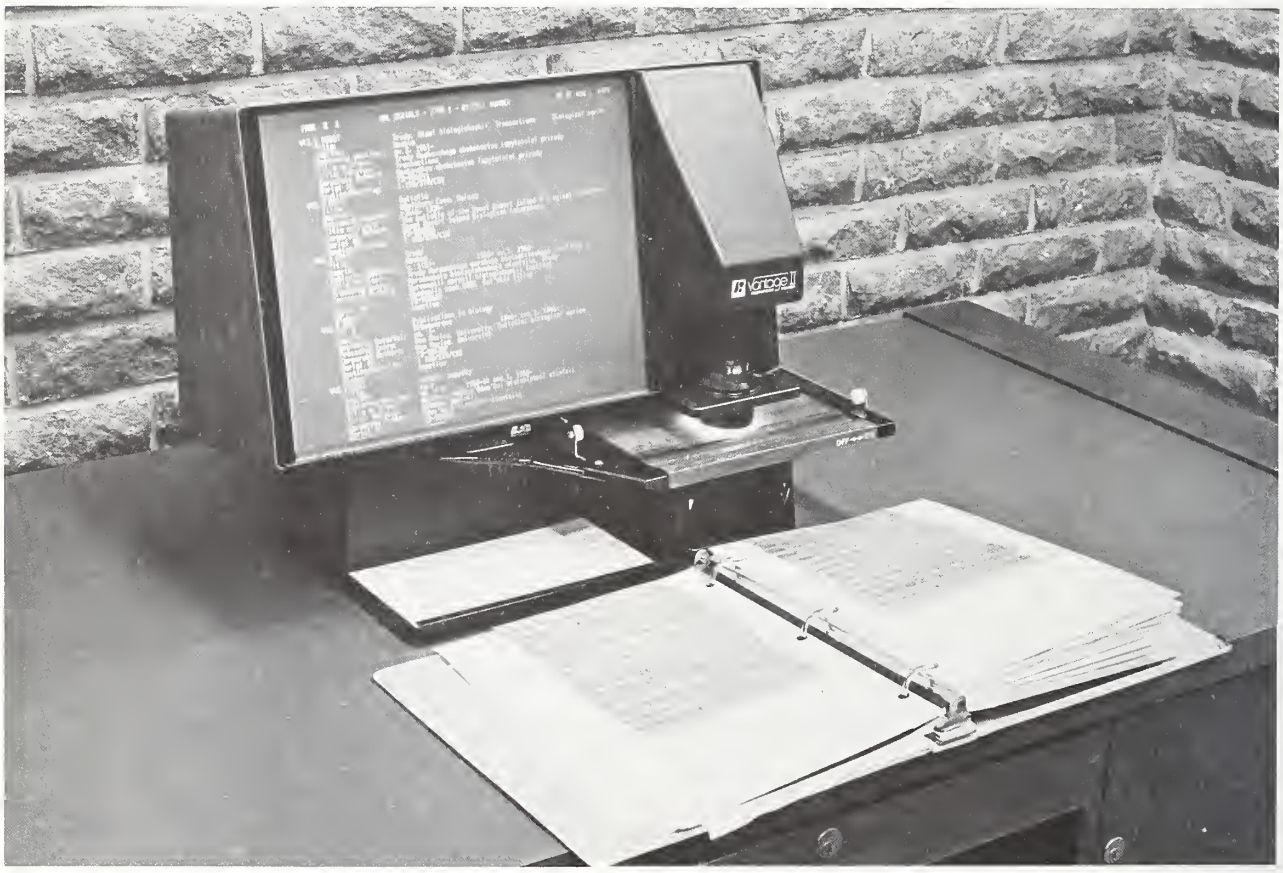
1. Martha E. Williams, "Education and Training for Online Use of Data Bases" (Paper presented at the EUSIDIC Conference, Graz, Austria, December 1, 1976), p. 4.

2. Charles L. Gilreath, *Cain Online User's Guide* (Washington, D.C.: U.S. Department of Agriculture, National Agricultural Library, 1976), p. 1-11.

3. Janet Egeland, "The Importance of User Education and Training in a Multi-Data Base Online Information Network," in *Information Utilities, Proceedings of the ASIS Meeting*, vol. 11 (Washington, D.C.: American Society for Information Science, 1974), p. 137.

4. Williams, *op. cit.*, p. 2.

5. Frederick W. Lancaster and G. B. Fayen, *Information Retrieval On-Line* (Los Angeles, Calif.: Melville Publishing Co., 1973), p. 320.



Computer output microfiche and reader
(Courtesy, National Agricultural Library)

THE USE OF THE AGRICOLA DATA BASE IN THE MODERN LAND-GRANT UNIVERSITY LIBRARY

by

Jerry V. Caswell*

Ever since their nineteenth century origins, the land-grant universities have been devoted to three basic functions: instruction, research, and extension. Instruction in practical subjects as well as the liberal arts signaled a breakaway from the classical tradition in education. Research was needed to apply the developing sciences of chemistry and biology to agricultural practice. Extension carried the fruits of knowledge to ordinary people.[1]

The research function has, perhaps, advanced more in the course of the twentieth century than the other two because of the expansion of scientific knowledge and the technological demands associated with it. Beginning with the Hatch Act of 1887, the federal government has continued to put its resources behind the research program of the land-grant universities. This has enabled many universities to develop first-rate facilities for highly sophisticated research.[2]

Unfortunately, for much of the twentieth century, library methodologies did not keep pace with the rate of scientific and technological development. While libraries increased in size and as the literature of the sciences grew, librarians continued to rely upon basically nineteenth century techniques for storing and disseminating knowledge. The development of abstracting and indexing services, useful as they were, still relied upon older techniques. *Chemical Abstracts* appeared in 1907; *Biological Abstracts* in 1927; and the *Bibliography of Agriculture* in 1942. Access to the serial literature was undoubtedly easier as a result of these tools, but became more and more time-consuming as the production of literature increased.

It was after the Second World War that we experienced the so-called "explosion" of scientific literature, which threatened to engulf scientists and librarians alike. In order to maintain control over the literature being published, abstracting and indexing services resorted to modern technology. By using computers to store and organize information, large amounts of material could be handled in an orderly way and segments could be retrieved for checking or re-editing. Out of this were born the data bases we use today.

All indexing and abstracting tools require author and subject indexes so that the user may find only those materials of interest. When these indexes are prepared so that they can be searched by, as well as stored on computer, a realm of new possibilities opened. The computer's capacity to retrieve and manipulate segments of the data bases gave us the searching ability upon which we currently rely.

The second major feature of computer search systems is the so-called on-line interactive aspect. In the beginning, computer searches were prepared ahead of time and then run in batches so that no one had any control over the output. To rectify an error or broaden or narrow a search required another complete search. By enabling the searcher to interact directly with the computer, as in modern systems, modifications could be made during the course of the search. This gave the user much tighter control over the output, since errors were corrected before printing took place.

It was in the summer of 1973, as most of us are aware, that the National Agricultural Library contracted with the Lockheed Missiles and Space Company to put on-line the CAIN data base upon which NAL had been working since the mid-1960's. Lockheed had developed a retrieval search system for the federal government called RECON which (in its revised format christened DIALOG) was used for translating the CAIN tapes into a searchable and retrievable system. Once the DIALOG system was operable, it became available to any institution or individual who wanted access to the information on the CAIN tapes.[3]

It was also in the summer of 1973 that the National Agricultural Library awarded a contract to Steenbock Memorial Library, the agricultural and life sciences library for the University of Wisconsin, to study the coverage of the forestry literature in nine abstracting and indexing tools. David K. Oyler, then Director of Steenbock Library, and his research associate, Michael McKay, selected a number of topics representative of current research interests and collected relevant citations from *Forestry Abstracts*, the Institute of Paper Chemistry's *Abstract Bulletin*, and seven other sources. Since CAIN had just become available on-line, it was decided to use it for searching the topics in place of the *Bibliography of Agriculture*.[4] A second purpose of the research project envisioned making free CAIN searches available to the University community in order to test the feasibility of the data base.

During the academic year 1973-1974, 137 searches were run in CAIN, the majority being for users in the College of Agricultural and Life Sciences. Because publicity was limited, response was slow at first. Later, as word spread about the service, use grew so dramatically that it began to impinge on the primary object of the research project. This necessitated scheduling appointments, a procedure to which the Library has adhered ever since that time. Entomologists, plant pathologists, and veterinary science researchers were the most frequent users of the service.[5]

At the conclusion of the test year the question arose: what to do with the service? It was clear by this time that it was of primary interest to graduate students and faculty who were involved in research projects or the preparation of dissertations. Yet there was no means by which the Library could underwrite the entire costs of the service. It was finally decided to proceed with the service on a partial cost recovery basis. The Library would charge back for computer connect time and off-line citation printing, while it absorbed the telephone fees, terminal rental, and labor costs. The responsibility for the service was given to the Reference Department and two of the librarians were trained in its operation. The

*Jerry V. Caswell, Reference Department, Steenbock Memorial Library, University of Wisconsin, Madison, Wisconsin

addition of computerized bibliographic retrieval services to the Reference Department was not only a natural extension of reference type services, but also actually increased the Department's ability to meet the needs of research personnel.[6]

Like many libraries, Steenbock had too small a staff to devote much of its time to the compilation of bibliographies for people involved in research. By offering computerized bibliographic service, the Library could actually become involved in a new area of activity.

As has happened in nearly every case, the imposition of charges resulted in an initial drop-off of use in the service. How long this would have continued is impossible to estimate. Early in the fall of 1974, the Deans of the College of Agriculture came to the Library for a demonstration and left promising support. What they did was set up a requirement that persons making applications for federal research funds under the Hatch and McIntire-Stennis Acts obtain a computer search as part of the application and screening process. They hoped to improve the quality of the Hatch applications by making the applicants more aware of the literature being published in their areas of interest. It is a commonplace to say that researchers simply are overwhelmed by the amount of time literature searching takes from other responsibilities. As a result, many have not kept up with the published research in their fields. The result of the action by the Deans was not only to upgrade the Hatch applications but also to acquaint researchers with a new and more efficient search process.[7]

For three years the Deans kept the requirement for Hatch applicants and for three years the use of the computerized search service grew correspondingly. As word of its value spread, non-Hatch users appeared as well. For the year 1974-1975, the Hatch searches numbered 74 out of a total of 193; the following year, 78 out of 293; and for fiscal year 1976-1977, they numbered 158 out of 546 searches.[8] Clearly, there was an overflow of interest from the Hatch applicants to others in the College.

The Library has used other means of arousing interest in the computerized searching. Information about the service and new data bases appears regularly in issues of the Library's *New Book List*. In addition, the searchers give demonstrations to seminars of graduate students and faculty and at the campus computer fair. Instructional guide sheets were developed at an early date to provide basic information to potential users.

AGRICOLA has been the most frequently used data base. During the first year of service, nearly all the searches were run in what was then called CAIN. In 1974-75, 134 out of the 193 searches were in CAIN; in 1975-76, 125 out of 293; and in 1976-77, 160 searches out of 546.[9] AGRICOLA indexes the 6,000 journals most relevant to agricultural research, is provided at modest cost, and gives up-to-date coverage of its fields.[10] These three reasons account for its popularity among researchers.

During the past three years the National Agricultural Library has funded several studies at land-grant universities for evaluating the utility of the data base. These reports, from Penn State, Colorado State, and the University of Minnesota, have examined many of the strengths and weaknesses of AGRICOLA in more detail than I can go into here.[11] In general, I think I can say that AGRICOLA is most effective with traditional types of agricultural research: soils, meat and animal science, forestry, plant diseases, and so on. For many of the new sciences, which may or may not be in agricultural schools depending upon the structure of the university, AGRICOLA is either inappropriate or needs to be supplemented by other data bases. Some of these areas are genetics, microbiology, the chemistry of food, and molecular biology.

As a result of their research at Steenbock Library, Oyler and McKay found that, in the area of forestry, AGRICOLA outranked every other abstracting and indexing tool, including such specialized ones as *Forestry Abstracts* and the *Abstract Journal* of the Institute of Paper Chemistry.[12] If this is the case in the other traditional agricultural areas where AGRICOLA is strong, it indicates that researchers will receive improved coverage in their areas of interest by using AGRICOLA in the place of many specialized indexing services. Thus AGRICOLA is not only economical but also of greater importance than many researchers might think at first glance. It would indeed be interesting to run some comparative studies in other subject areas to see if this hypothesis could be borne out.[13]

The ability to retrieve effectively from a data base is just as important as the material which is entered into it. Since subject searching in AGRICOLA is limited to title words plus enrichment terms and some general category codes, it is not always possible to retrieve all the citations on a requested subject. A formal vocabulary of descriptors would certainly enhance subject searching in AGRICOLA as would more detailed subject category codes. It would be invaluable, for example, to limit the results of certain searches just to animal, plant, or insect studies. It would also be useful to be able to group all studies of domestic animals, ungulates, or poultry together. Such refinements would require more indexing efforts initially but would result in greater utility to the user.

Both in terms of speed and the ability to manipulate numbers of terms, computerized searching has more appeal than manual searching. Yet, there is an interesting spinoff from the AGRICOLA searches. More people appear to be using the *Bibliography of Agriculture* than ever before. I think we may attribute some of this interest to the AGRICOLA computer searches which have now been widely disseminated about the campus.

AGRICOLA has undoubtedly enhanced the research efforts of the land-grant institutions where it is used.

Besides reaching new users, it has taken a burden off of the shoulders of researchers by helping them keep abreast of developments in a timely fashion. Due to its provision of such data bases, the modern agricultural and life sciences library has joined the mainstream of the twentieth century technological development. It can now provide a sophisticated form of reference service--the individually tailored bibliography--which it could not hope to provide before. This resource greatly increases the Library's ability

to serve the modern research needs of the College of Agriculture to which it is attached.

REFERENCES

1. Charles E. Kellogg and David C. Knapp, *The College of Agriculture: Science in the Public Service* (New York: McGraw Hill, 1966), p. 4; American Association of Land-Grant Colleges and Universities, *Land-Grant Fact Book* (Washington, D.C., 1962?), pp. 2-3.

2. Kellogg and Knapp, *College of Agriculture*, pp. 12-13; AALGCU, *Land-Grant Fact Book*, pp. 29-30.

3. Hermes D. Kreilkamp, "The National Agricultural Library's Data Base: AGRICOLA," *College and Research Libraries* 38 (1977): 299.

4. David K. Oyler and Michael W. McKay, *The Impact of the Forest and Forest Industry on the Environment: A Study of Bibliographic Coverage* (Madison, Wis.: University of Wisconsin, 1975), p. 7.

5. David K. Oyler and Michael W. McKay, *An Analysis of CAIN On-line, University of Wisconsin-Madison* (Madison, Wis.: University of Wisconsin, 1975), p. 3, 5-6.

6. Oyler and McKay, *Analysis of CAIN On-line*, pp. 8, 14.

7. Jerry V. Caswell and David K. Oyler, "Specialized Data Base Utilization," *Agricultural Libraries Information Notes* 2, #3 (1976): 2.

8. Jerry V. Caswell, *Report(s) to the Research Division on Hatch and McIntire-Stennis Applicants* (Madison, Wis.: Steenbock Memorial Library, 1975-77).

9. Jerry V. Caswell, *COMBIDS Annual Report(s)* (Madison, Wis.: Steenbock Memorial Library, 1975-77).

10. Charles L. Gilreath, *CAIN Online Users Guide* (Beltsville, Md.: National Agricultural Library, 1976), pp. II-9.

11. Curtis L. Gifford, *CAIN On-line Testing and Assistance at Colorado State University* (Fort Collins, Col.: Colorado State Universities Library, 1976); Phyllis M. Reich and Fred E. Hearth, *Evaluation of CAIN On-line; Final Report to National Agricultural Library* (St. Paul, Minn.: University of Minnesota, 1975); Vladimir Micuda, Keith Roe, and Robert Seeds, *CAIN On-line Testing and Assistance, 1974-75* (University Park, Penn.: Pennsylvania State University, 1975).

12. Oyler and McKay, *Impact of the Forest and Forest Industry*, pp. 9-11.

13. An analysis of the gypsy moth literature carried out at the University of Wisconsin indicates that, while AGRICOLA is the most comprehensive file in the short run, CAB Abstracts may equal it over the long term. Mary L. Cittadino, Ronald L. Giese, and Jerry V. Caswell, "A Comparison of Three Computer-Based Bibliographic Systems for Scientific Literature," *Bioscience* 27, #11 (1977).

PUDOC AS USER OF THE NAL DATA BASES

by

H. C. Molster*

Introduction on Pudoc

Before giving an exposé of the use of CAIN and related services by Pudoc and sharing with you some of my concerns for the future, I should give you a short introduction on the function and structure of my organization.

The Centre for Agricultural Publishing and Documentation (Pudoc for short), which was founded in September 1957 in Wageningen,[1] is a legal entity under the Dutch Ministry of Agriculture and Fisheries and is directly responsible to the Directorate for Agricultural Research. It was created to service the informational needs of the agricultural community in the Netherlands with emphasis on information for agricultural research.

Most Dutch agricultural research is concentrated in and around the town of Wageningen in the eastern part of the country. The library of the Agricultural University, also located in this town, is the major agricultural research center in the country. Even though Pudoc uses this library quite extensively and works closely with the faculty of the University, it has no organizational ties with the Agricultural University or its library. This may soon change, however, since plans for formalizing cooperation are under way. In 1980, we hope to be at least housed in the same new building.

The division of work between Pudoc and the University Library has generally been quite clear: the information function of the library is restricted to making its own collection accessible, helping people to find relevant material that is available locally as well as in the reference room.

Pudoc, on the other hand, does literature searches, some specialized bibliographies, and literature reviews. In addition, Pudoc operates a manual current awareness service for some 40 clients. Incoming material is also screened for inclusion in the AGRIS system since Pudoc is the Dutch AGRIS in-put center (as well as the AGRIS liaison office). A computerized current awareness service was started in 1973 and the use of on-line information retrieval in 1974. Details on both will be given in the next section.

Pudoc's Department of Documentation also publishes three secondary journals:

Landbouwdocumentatie (*Agricultural Documentation*), a weekly bulletin in Dutch with annotated titles of general interest to agricultural scientists, extension officers, agribusiness, etc. in the Netherlands;

Pudoc Bulletin, a quarterly bibliographic list of publications by Dutch scientists;

Agricultural Aspects of the Common Market, a monthly list of publications related to agricultural affairs of the European Common Market.

From the annotated titles of *Agricultural Documentation*, a card file is maintained which serves as a first entry for many incoming questions, but its highly selective coverage (only between 7,000 and 8,000 titles per year) does not make it suitable for wider use as a documentation tool.

All bibliographies made on request and various other bibliographic material are filed in a UDC system. Apart from these activities, Pudoc does not make or maintain any other documentation although it relies heavily on the documentation systems of others. In addition to helping people find relevant literature for their research, Pudoc is also involved in the production of primary documents. The Publishing Department produces about 35 titles per year. It also offers editorial services to several journals of learned societies as well as facilities for English translation and revision. For all its functions, Pudoc is still just a small organization with a total staff of around 40 people. Computer services may enable us to do more for more people without an appreciable increase in staff.

Pudoc's Use of Cain

Pudoc's use of the *Bibliography of Agriculture* has been less extensive than that of the journals of the Commonwealth Agricultural Bureaux (CAB). We were quite surprised to learn that only six out of 102 libraries of institutions related to agricultural research in the Netherlands subscribed to the *Bibliography of Agriculture* in 1971, with 20 subscriptions to *Chemical Abstracts*, 16 to *Veterinary Bulletin*, and 15 to *Biological Abstracts* and *Horticultural Abstracts* to mention but a few. The fact that CAB has good specialty journals for most of the specialized fields of the research institutes probably explains the limited usage of CAIN's printed version.

When Pudoc first considered the use of machine readable files for improving its information services, the existence of the CAIN data base played an essential role. All SDI profiles that were run in the initial feasibility study for the use of computerized service were run on CAIN; roughly half of them were also on other data bases. CAIN (AGRICOLA) is still our main data base both for our SDI and on-line services: for SDI, we process the CAIN tapes under our own control; for retrospective searching we use the Lockheed and SDC facilities.

THE SDI SERVICE - *The Initial Feasibility Study*

In 1971, a proposal was drafted for an evaluation of CAIN and Food Science and Technology Abstracts (FSTA) for current awareness, using a computer retrieval program developed at IWIS-TNO, a Dutch semi-governmental computer center. The study specifically compared the search results with those of Pudoc's manual current awareness service. Funding for the

*H. C. Molster, The Centre for Agricultural Publishing and Documentation (PUDOC).

study was obtained from NOBIN, the Dutch organization for the advancement of informational services.

The first profiles were run on *Chemical Abstracts* in May 1972 and the first FSTA and CAIN tapes were processed in September. In April of 1973, I assumed the position of project leader at Pudoc.

Gradually the number of profiles was increased from the original nine in 1972 to the envisaged 100 in the spring of 1974. Participants in the experiment obtained the output free with the proviso that they should evaluate each "hit" on its relevancy and novelty. The participants were not chosen on the basis of any formal sampling model but because their backgrounds seemed as varied as the total set of potential users. Twenty-five of the profiles were specifically chosen for comparison with the manual current awareness service.

The project came to a close after the March 1975 tapes of CAIN, FSTA, *Chemical Abstracts* and *Biological Abstracts*. *AGRINDEX* and CAB were run somewhat longer under this scheme.[2]

The final report on this study[3] concludes that the project demonstrated the technical feasibility of a current awareness service for agricultural scientists based on existing computer-readable files and retrieval systems.

The major indicator for user acceptance of the service was the percentage of participants who continued their subscription after the experimental period was over. It turned out that 60 percent of the participants wanted to continue which was considered quite a success.[4]

The major results of the output evaluation by the users are summarized in Tables 1 through 5.[5]

Table 1. Relevancy judgement in percentages of all titles found.

Database	++	+	±	-	±/++,+,±	++/++,+
CAIN	14	28	20	38	0.30	0.33
FSTA	6	35	21	38	0.28	0.15
CAC	5	23	17	55	0.35	0.18
BA	16	13	14	57	0.25	0.55
Medlars	13	39	16	36	0.22	0.25

The computations were made by averaging the averages per profile. If all titles are considered as one set (which overvalues the results of profiles with a high volume of output) the results are generally somewhat lower: profiles with a relatively high volume of output had thus generally somewhat lower relevancy ratios than profiles with lower output.

Table 2. Frequency distribution for the relevancy categories 'relevant' and 'perhaps relevant' together. (To make comparisons possible, they have been converted to percentages.)

Database	0-25%	26-50%	51-75%	76-100%
CAIN	3	34	34	29
FSTA	0	23	45	32
CAC	12	47	26	15
BA	20	50	20	10
Medlars	0	22	56	22

Table 3. Average number of titles per year per profile.

Database	++	+	±	-	Total
CAIN	22	44	35	59	160
FSTA	10	66	42	70	184
CAC	10	84	48	199	355
BA	30	95	86	515	742
Medlars	22	84	21	52	178

The differences between the totals of the first four columns and the fifth column are due to rounding errors. These numbers have been calculated by multiplying the average number of hits per run by the number of runs per year. CAIN, FSTA and Medlars have 12 issues per year, Chemical Abstracts 24 and BA-Previews 36.

Table 4. New titles as percentage of all 'relevant' and 'perhaps relevant' titles.

Database	n/++	n/++,+	n/++,+,±
CAIN	33	60	59
FSTA	25	58	58
CAC	67	71	67
BA	56	62	58
Medlars	36	46	50

'n/++' is the quotient of all new titles over all 'very relevant' titles, expressed as percentage. 'n/++,+' the same for all 'very relevant' and 'relevant' titles and 'n/++,+,±' the same for all not irrelevant titles.

Table 5. Average number of relevant new titles per profile per year.

Database	++n	+n	±n	Total
CAIN	11	26	17	54
FSTA	2	35	26	63
CAC	5	58	31	94
BA	19	72	54	145
Medlars	12	54	11	77

The Comparison with the Manual Current Awareness Service

Before any figures on this part of the study are given, it should be made clear that it concerns a comparison between services which differ in at least three ways: as retrieval medium there was, on the one hand, the

computer retrieval program plus the profiles constructed by the search formulator, while, on the other, there was the information specialist with a general, non-formal picture of the client's interests; the search object was, on the one hand, bibliographic data files of secondary material and, on the other, primary documents; the collection of primary documents screened for inclusion in the bibliographic files was different from the collection screened for the manual current awareness service. Nevertheless, the comparison seems valid because it compares two options between which Pudoc and its clients can in reality choose.

The final results can be expressed in a few dry figures. When all relevant titles of the computer and manual service are put together, with duplicates removed, the manual service found 30% of all titles and the computer service 80%.[6] Of all titles found manually, only 33% were also found by the computer service, and of those found by the computer service, only 12.5% were found manually.

These averages illustrate not only that the two services generally complement each other but also that when a choice is necessary, the computer service will give one more than the manual service. These figures are only averages; behind them is a large variety of hidden individual outcomes. The former show clearly that for each individual information need, a separate decision has to be made. There are a few special circumstances where the manual service has a clear advantage:

necessity for fast delivery of the original document (As soon as the manual searcher spots a relevant article, he can make a photocopy and send it to the client. The photocopy may be at the client's desk earlier than the journal is available in the reading room);

areas of interest, which are well-covered by the library where the manual service is based, but less by the international documentation systems;

areas for which no efficient profile for computer retrieval can be written;

areas which, if covered in an article, are often not specifically mentioned in the title or index terms (especially on methodology);

selection of qualitative criteria on which no computer retrieval is possible such as "important articles."

The manual service was generally a little bit *faster* than the computer service in supplying the title of a particular document but the average differences were small. CAIN alerted a client to a title on the average two months later than the manual service. *Chemical Abstracts* was even faster by two months. FSTA was five months slower and *Biological Abstracts* eight months. A major disadvantage of our manual current awareness service is its small maximum volume and high labour intensiveness. Limited capacity of the human brain makes it such that any one information specialist can only differentiate between 25-50 profiles. For the time being, we will continue to offer the service for a maximum of around 50 profiles.

Development and Present Status

After the feasibility study was over, the Ministry of Agriculture agreed to assume the financial responsibility for the service, with the condition that a subscription fee should be charged to cover at least the computer cost. All clients working at institutions of the Ministry would be allowed a rebate of 50 percent, as budgetary provisions at the various research establishments might not yet be sufficient to accommodate a large number of subscriptions. The number of paid subscriptions rose from 116 in 1975 to 155 in 1976 and 235 in September 1977. These are subscriptions per subject, independent of the number of data bases on which each profile is run. They are also independent of the number of copies produced of the same profile for different people. The number of data bases per profile is at the moment 1.8. The increase of the number of profiles is mainly dependent on the manpower available to visit potential clients and convince them of the advantages of the service. The potential number of end users in the Netherlands is around 1,250.

At the moment Pudoc processes four data bases under its own control: CAIN with 107 profiles; CAB with 124; AGRINDEX with 79, and FTSA with 32. In addition, 54 are run on *Chemical Abstracts* administered by a different organization, but with the same computer system, and 25 on *Biological Abstracts* at the Biological Information Service, University of Nottingham in the United Kingdom.

The fact that CAB is used more than CAIN is mainly due to the preferences of two large clients: the Faculty of Veterinary Medicine prefers CAB and has 24 profiles on CAB alone, and the Ministry of Agriculture, Food and Fisheries of the United Kingdom invests generally only in one data base per profile, which is CAB for 36 and CAIN for 12 cases.

Pudoc does not plan to add any new data bases to its service, although *Biological Abstracts* might be a possibility. Postal services to and from the United Kingdom are far from optimal; the longer processing chain increases the number of error sources. Thus, if we could afford it, we would rather do it ourselves, but a hundred profiles seem to be an economic minimum and that number is not yet in sight.

The software[8] has been continuously improved. It is a full-text searching program with the normal Boolean logic facilities and truncation at both ends. The IGNORE facility has proven to be very useful.

Pudoc also has an on-line interactive updating facility. As far as costs are concerned, there is some competition from Lockheed's SDI service, but for the time being our clients seem willing to pay a little extra for the library format cards they get as output from us compared to the paper output from Lockheed, which still needs retyping or cutting and pasting before filing.

At the end of this resumé you may ask: "and what about CAIN?" CAIN has in fact become so much a standard feature of our life that one forgets to appreciate its availability. It has long been the only data base available in computer readable form that covered the scientific agricultural literature of the world. It is still by far the cheapest. Its shortcomings must

be well-known to you; I do not want to dwell upon them now. It is not so easy to retrieve as many relevant titles as possible using the title and a limited set of category codes alone, but it works quite well. We appreciate the fast increase in the percentage of enriched titles. The ideal data base for information retrieval should, in my opinion, contain a more detailed set of hierarchical codes for concepts, organisms, chemicals, and geographical areas. The concept of AGRINDEX seems promising in this respect.

The Use of Computer Readable Files for Retrospective Searching

The availability of the full CAIN file on-line for retrospective searching added a major new item for agricultural information retrieval. The first connection was made in May 1974. In that year, Pudoc used a total of 118 hours of connect time of which 83 were CAIN. Subsequent use of our connect time with Lockheed, SDC, and the European Space Agency in Frascati, Italy, is summarized in Table 6. The figures may be somewhat disconcerting as it shows a fast levelling off of our total use and an appreciable drop in CAIN use over the first four years. The number of bibliographic requests increased considerably, however, especially during the present year. In 1976, Pudoc processed 150 external requests whereas over 200 are expected to be received during the present year. In 1975, 207 questions were processed, 44 of which were for a special study for the European Commission on the coverage of veterinary medicine by the various data bases. The time required per question was 30 minutes in 1975 and 50 minutes in 1976. No figures are yet available for 1977. Time per question is increasing due to the increase in the number of data bases which are relevant to our area of specialization.

Table 6. Computer connect time for on-line bibliographic searching used by Pudoc, 1974-1977.

Database	1974[a]	1975	1976	1977[b]
CAIN	83 (71%)	108[c] (53%)	87 (44%)	59 (27%)
CA	21 (18%)	23 (11%)	24 (12%)	32 (15%)
Biosis	-	22 (11%)	27 (14%)	32 (15%)
CAB	-	-	-	29 (14%)
Other	13 (11%)	51 (25%)	61 (31%)	62 (29%)
Total	117(100%)	204(100%)	199(100%)	214(100%)

[a] first searches in May. On year basis the figures would be: CAIN, 111; CA, 28; Other, 18; Total, 157.

[b] extrapolated from actual use until July 31st.

[c] including 44 questions for a special project of the European Commission.

The decrease in relative use of CAIN may be explained by the continuous increase in available relevant data bases. There is a trend towards more requests from people involved in basic research which explains the increased use of *Chemical Abstracts* and *Biological Abstracts*. The availability of CAB since the beginning of this year has made some dents in our CAIN use, even though we searched both data bases in a majority of cases. The main disadvantage of CAB is the lack of integration of the constituent parts and the fact that the search terms are limited to title and keywords only, giving too many false drops. The high costs of CAB compared to CAIN are, of course, also a major consideration in the decision whether or not to use CAB.

The use of on-line retrospective searching has meant the addition of a completely new service, even though Pudoc actually had offered the same type of output (i.e. typed bibliographic lists). It is still the philosophy that the tools to be used for answering a particular question be discussed between the requestor and the information specialist with the latter having the final word. In some cases manual searching may prove more efficient. Ideally, Pudoc would like to offer every client a typed bibliography as the result of the integration of all computer and manual efforts. However, this has proven to be generally too labour-intensive. The decision to use either manual or computer methods or both is still taken on an *ad hoc* basis. With the introduction of computer searching, Pudoc had to abandon its principle of free bibliographic services. For each question that is considered "serious" (defined as expected to take more than four hours search time if processed manually), Pudoc charges a flat fee of $100. Most computer output is now sent to the clients as it comes in from the processor, after some evaluation on the part of the search formulator.

The use of on-line retrospective searching has thus changed Pudoc's work pattern considerably. It also has made tailor-made bibliographic help available to a much larger number of people. Increasing the number of present users is impeded by much the same factors as are valid for the SDI service: limitations of staff at Pudoc, the traditional attitudes of scientists who are not easily convinced of the efficiency of external information services, and problems with budgeting. Thus, in the near future, those of us at Pudoc expect a steady but not an especially fast growth in the use of our services.

Factual Information and Document Delivery

For input, the main long term problem is, how can we more directly satisfy the scientist's needs for information instead of just giving him bibliographic references? This concerns the whole field of information analysis and data banks which will certainly need a great deal of attention in the future.

For completeness I will mention another problem: the need for improved efficiency in supplying documents. The NAL system has, especially in the United States, the big advantage of a library-based data base which guarantees the availability of the original document from NAL. Even though the NAL call number is on every CAIN record, it is still difficult to get a Dutch librarian to request it by Telex from NAL if it is not available in Wageningen. The system of international loans has not yet been accepted as a normal facility.

For those data bases where the producer does not offer a document delivery service, it is often less clear how the original can be obtained. As the international data bases bring more information to the user, library staff need to adapt more to the regular use of the various national and international lending facilities.

Data Base Multiplicity

The major problem of input for the near future is the multiplicity of data bases. Their proliferation and the availability through vendors of on-line services have made the selection and combination of data bases quite an art. It is becoming more difficult to decide on the optimum data base mix in terms of coverage and costs.

The agricultural sciences present a special problem because three data bases are available on-line, all claiming to cover the world literature in agricultural sciences. Besides CAIN, there is the service from the Commonwealth Agricultural Bureaux (CAB) and from AGRIS of FAO. The CAB service is available through Lockheed; AGRIS is, to my knowledge, not yet available on-line in the United States; in Europe, an experimental service has just started from the International Atomic Energy Agency in Vienna, and the European Commission has the data base on-line for internal use. The differences between the three data bases are probably well-known to you; they are summarized in Table 7.

there are still some major areas of agricultural sciences for which CAB coverage is weak, such as poultry science and agricultural engineering. However, CAB is certainly working hard to improve itself in all these areas.

No comparative figures are available yet on the delay of CAB relative to CAIN and AGRIS; our impression is, however, that CAB will on the average be no more than a few months behind the others. CAB is thus certainly tending towards a comprehensive agricultural data base. It incorporates a high intellectual effort at the input stage which is a big advantage and the selectivity which some find an advantage and some do not.

In the politics of information, CAB seems to be trying to make its bibliographic services self-sufficient and politically independent. Although controlled at present by the Commonwealth and, in that sense multinational, a more intensive international cooperation seems desirable.

For exhaustive coverage, we will inevitably have to

Table 7. Differences between AGRIS, CAB and CAIN

	AGRIS-I	CAB	CAIN
Coverage	Some major countries not yet fully covered; all areas of agriculture within scope.	Selective; some areas not yet fully covered: animal husbandry, agricultural engineering. Includes some areas of human medicine.	All relevant US literature. World literature selected according to US needs.
Organization	Decentralized, but with high degree of standardization.	Decentralized within UK. Some cooperation from selected institutions in other countries. Standardization planned but not yet far developed.	Centralized at NAL.
No. of unique documents per year	Approaching 100,000	Approx. 115,000. Not exactly known because of internal overlap (including some material for human medicine).	Approx. 130,000
Coding	85 subject codes 420 commodity codes 560 geographical codes	Large differences between bureaux. No standardized system for all of CAB. Generally quite developed.	70 subject codes 250 geographic codes
Keywords	No	Yes. Practices differ between bureaux.	No (only for FNIC and AG-Econ)
Abstracts	No	Yes (with some exceptions, notably Index Vet.)	No (except for FNIC and AG-Econ)

Since former directors of both CAB and NAL have served on the original study group that recommended the AGRIS concept to FAO, one would hope that some convergence of the three systems could be attained. However, times have changed considerably since the inception of AGRIS, both technically and politically. Originally it was probably hoped that the CAB services could be part of "AGRIS level two" as part of a network of specialized information centers, each responsible for the detailed accessibility of the literature within their scope, whereas AGRIS level one would offer a central service with limited intellectual effort for input, to facilitate fast dissemination. However, CAB is now offering a complete service covering all agricultural sciences and including abstracts. The CAB data base still suffers severely from lack of standardization and central control over the various input units. CAB cannot yet live up to its claim of complete coverage for all the relevant literature;

search both CAIN and CAB; we will have to put up with the resulting overlap and the extra cost as long as no cooperative solution is in sight. As both data bases have different primary missions and finances are not yet a constraint for the producers, the duplication seems unavoidable for the time being.

The situation of AGRIS is quite different. The major decision on the continuation of AGRIS will be taken by the FAO Conference in November, 1977. As the FAO Council has not advised negatively and the UNESCO evaluation report[9] recommends continuation, it would seem that the Conference will approve continued funding for AGRIS level one. Especially for the developing countries, AGRIS has already proved a major stimulus for building their own information infrastructure. This reason alone warrants continuation of the system. It has also been proven that description of national documents by national input centers gives

much better coverage than input by a central body, expecially for "gray" or "non-conventional" literature. The main weakness of the system is that comprehensive coverage of all agricultural literature within scope is only attained if all countries producing such literature participate.

Through personal experience and involvement, I have become a strong supporter of international cooperation in agricultural documentation, especially with a seemingly sound concept like AGRIS. The main weakness of the pilot project was its lacunes through lack of cooperation of some major countries, including the United States. One could not expect a pilot phase only operational for three years to be fully comprehensive. Thus it is unfair to judge AGRIS on that weakness. AGRIS on-line is, at present, a service of very limited usefulness, as it has only just over 100,000 records even though it contains many records not to be found in CAIN or CAB. As a monthly SDI service, we recommend it as an additional data base to those who require exhaustive coverage and for subjects in which we know AGRIS is relatively strong (i.e. literature from developing countries and from Japan). It is my sincere hope that the United States understands the importance of including all relevant American literature in the AGRIS system. All necessary titles are already put into CAIN but they need extra indexing to meet AGRIS standards. I understand the difficulties involved in changing some of the input procedures from CAIN to AGRIS format, especially since the CAIN system was first in operation and performs very well for national purposes. However, since the AGRIS system is based on newer international standards and its records can be more effectively searched by the user, it would be a great benefit to agricultural science all over the world if the United States were willing to cooperate fully.

However much I appreciate the different missions of CAIN, CAB, and AGRIS, I sincerely hope that we will deal with one comprehensive agricultural data base to serve the bibliographic needs of agricultural scientists in the not too distant future.

Finance and Pricing Policy

On the other side of the coin, there are the interrelated problems of efficient use of the service, finance, and pricing policy. In our experience, scientists do not readily accept tailor-made information from others, even if it is provided free. Replacement of a major part of the literature selection traditionally done by the scientist himself by an external service involves a change of attitude. Advertisements and articles about these services have brought us very few clients. However, almost every personal visit to a scientist has paid off in a new SDI subscription or the request for a retrospective search at least if budgetary provisions were available for external information services. Librarians at research establishments have generally not been involved in the dissemination of our service. Their active involvement with individual information needs is generally limited to the information contained in their own library.

In the previous paragraph I have assumed that funding for the service was no problem. This is not true in actuality, however. Only for the agricultural university people is the service virtually free because they have created a central budget for external information services, which relieves the departmental budgets of these costs. Only the approval of the head of the department is required.

The lack of such budgetary provisions in the research institutes of the Ministry of Agriculture seems to have caused a serious delay in the use of our services as compared to those of the University. Since it is generally cheaper to obtain information on research already carried out than it is to set up new experiments, the cost of information should not interfere with obtaining information. Thus the need for improvement of funding for information services in our country is evident.

Our Director of Agricultural Research at the Ministry of Agriculture has expressed the view that (especially in hard economic times where it becomes more and more difficult to obtain research funds) before experiments are set up, extra attention should be given to the use of information services in solving problems. However, this statement has not yet led to a more innovative policy in information funding. This requires a long process of diffusion before new policies become generally accepted. Only some restructuring of our research organization would make it possible to use the policy instruments in a more efficient way. Up to now, no extra staff and little extra money have been provided for information services from the central research budget. Only a redistribution of budgets within the research units will make more money available for information services. The outcome can already be guessed: there will be no redistribution at all, since hardly any researcher will divert research money to information.

The question of funding and pricing information services also involves some more theoretical issues which I want to touch upon briefly. One external solution for funding is to finance it at the source and provide all information services free to those in need. This would be the most efficient way of making information services available to all as fast as possible.

However, at present, this solution is not politically feasible in our country. At least one feature makes such information services as ours different from traditional library work: for individual literature searches the direct costs for the fulfillment of each individual request are high relative to the overhead which is much less for loan requests.

It could rightly be argued that each client should pay at least for the costs that he has incurred himself. Recently, a Dutch commission on pricing for information services reached this very conclusion: the client should pay his own costs and the government should finance the rest directly. We will have to wait and see whether the government will accept this recommendation and the consequences.

The other extreme solution for the financing of information services is to let the user bear the full cost including staffing. Such a solution would be in line with ideas in government circles of full-project financing, whereby all costs including invisibles and overheads, are charged to the project and the project

is funded for this total. Such a policy would make real costs much more transparent and would make it easier to compare such costs among institutions. However, until this way of financing research projects is implemented, it would be self-defeating to be the first governmental institution that requires full-cost payment for its services.

Finally a few remarks on price structure. In our country, it has been the general practice to let the client pay as nearly as possible the costs incurred in his behalf, sometimes with an overhead charge. Although this may seem fair and simple, it can also be argued that it is unfair and difficult. First of all, with such a system the client never knows in advance how much a search will cost him unless he sets a limit. Secondly, he will be lucky if a data base exists that closely suits his particular needs, but will be penalized if his query happens to be in a fringe area of quite a few data bases and thus requires the use of all these data bases. This also applies to the number of search terms needed: if a data base is well-structured for a particular query, only a few search terms will be necessary but, if it is not, you may have to write out, for example, lists of organisms, if there is no group-code for them. Finally there is the dependence on the searcher who may be experienced or green, and who may have a perceptive active mood or an off-day. After considering all of these arguments, Pudoc decided from the start to ask a flat fee for each question independent of costs, the number of data bases used, and whether any manual searching was done. Our philosophy is that what the client needs is a satisfying answer for a price known to him in advance. It is of no concern to him how the answer is obtained. We may actually do a more efficient job by not using the computer at all, if we happen to know some other source that should be sufficient or even superior. But why should the client pay less in such a case? After long discussions, the national committee on pricing policies has agreed to this point of view. We hope, of course, that it will be accepted by the Dutch government!

For SDI services, we have not yet gone so far as one flat fee. At the moment our price is dependent on the number of data bases searched only, not on the number of search terms or on the output. We hope to change this policy in 1978 to a structure whereby the price depends only on the number of records searched per year, with four number classes. A combined service will then be offered of CAIN, CAB, AGRIS, FSTA, and CAS.

NOTES

1. The Documentation Department was established in 1951 at the Ministry of Agriculture and Fisheries in the Hague.

2. It was the original intention to include also the AGRINDEX and CAB files in the study, but they were both available too late to do so. Nevertheless we were able to run the first AGRIS year, 1975, and the veterinary part of CAB in the same year on money from the original grant. In 1973, the MacMillan version of the CAIN file was processed alongside CAIN.

3. H. C. Molster, 1977: Het ALADIN-project. De computer als hulpmiddel bij de literatuurattendering ten behoeve van de landbouwwetenschap. (The "ALADIN project." The computer as aid to the current awareness service for agricultural sciences.) In Dutch, 134 pp. Centre for Agricultural Publishing and Documentation, Wageningen, the Netherlands.

4. This percentage applied to continuations for CAIN. For FSTA it was 54% and for CA 45%.

5. The total number of evaluated records on which the results are based were:

	Relevancy	Novelty
CAIN	12,425 (59)	9,787 (53)
FSTA	4,224 (16)	3,668 (13)
CAC	13,776 (22)	8,780 (16)
BA	3,566 (8)	3,566 (8)
Medlars	1,172 (6)	1,172 (6)

The evaluation categories for relevancy were:
++ : very relevant for my subject of interest
+ : relevant to my subject of interest
± : cannot judge from the title, whether this title is relevant or not. Also: of borderline interest, or of interest to me, but not directly related to the specific subject.
- : irrelevant

News value was defined as "have not yet seen this title before." Participants were specifically instructed to add the "n" for "new" also when they expected the title soon in a journal they subscribed to or that was regularly circulated, but in which they had not yet actually seen it.

6. If one takes the average of the averages per profile the figures are 35% manual and 76% computer.

7. An additional 100 profiles are run both on CAIN and on CAB for the CAB/CAIN comparison project, financed jointly by the British Library and Pudoc.

8. The package is called FLIRT, which stands for Fast Literature Information Retrieval Tool. It is especially written for a Control Data-Cyber machine. It runs presently on a Cyber-72 under NOS/BE level 434 with 128 k word memory (60 bits/word), 10 peripheral processors with 4k words and 1,000 megabytes disk memory.

9. Badran, O. A., J. Haman, F. W. Lancaster and J. Martyn. Report on the independent appraisal of AGRIS. UNISIST report nr. SC/77/ws/20 UNESCO, Paris, April 1977.
ISBN 92-5-100309-2.

RESEARCHER'S EVALUATION OF ON-LINE DATA BASES

by

Ronald H. Goodwin*

It is challenging to be called upon to present the researcher's viewpoint of data bases available through the facilities of the National Agricultural Library.

While I have mainly used the batch system CAIN output together with other data bases provided by the ARS Current Awareness Literature Service, I have studied the on-line AGRICOLA service now offered and consulted with users. Together these systems provide a vastly increased bibliographic searching power which I believe has greatly benefited my research investigations.

Before using data base information, I prided myself on being a regular and avid follower of ISI's *Current Contents of Life Sciences and Agricultural Sciences*. Ever since my graduate school days, I have used these publications as my "standard library service" especially when I was employed by the Australian Government at a rural experiment station in New South Wales.

I was trained as an invertebrate pathologist, although I have been classified by USDA Agricultural Research Service (ARS) as a research entomologist. For the last eight years, I have conducted studies in insect cell and tissue culture with the practical aim of mass culturing certain insect viruses in controlled fermentors for field application against forest, agricultural, and public health insect pests. As you may realize, this represents work in one of our priority research areas; namely that of Biological Control and Integrated Pest Management. Insect cell culture study also provides a powerful new basic research tool for insect physiology as well as insect pathology. The current awareness profiles that I have developed are in invertebrate cell and tissue culture. I have also run retrospective searches in this area for a review covering the 1972-1976 period.

Invertebrate pathology encompasses both the microbial world and the physiological and physiopathological study of the one group of animals, namely the insects and other invertebrates, which dwarfs other groups in its taxonomic diversity. Ninety-seven percent of the animal species on earth are invertebrates, and 75 percent of invertebrates are insects.

Perhaps unconsciously recognizing the information problem facing me as a student of invertebrate pathology, I quite early attempted to broaden my bibliographic coverage in order to include the relevant but highly disparate and scattered published tools that I would need to review constantly in order to become and remain competent in my field.

Continued weekly in-depth survey of the issues of *Current Contents* was and still remains the only satisfactory recourse to meet this need. What I have done in this instance was to resist definition of my background reading and continuously try to maintain an open-minded curiosity about new observations and new ideas at the fringes of relevant subject categories, allowing these to feed new directions in the conceptualization of my specific research problems.

I think that there may be a narrowing tendency in overdependence on closely defined current awareness data base profiles, even with continuous periodic revisions. This tendency may become more of a concern as researchers come to depend more and more on data bases as their primary bibliographic tools. In addition, as keywords may change from those initially entered on a profile (including taxonomic reclassifications, chemical name changes, or regrouping of any subject categories into new groupings), one must periodically return to journal and/or *Current Contents* perusal in order to pick up changes which may occur and which may result in citation call down failure in a formerly satisfactory profile.

From what information I have it does seem, at least among ARS researchers, that the Current Awareness Literature Service has become so widely used that it overshadows most other literature sources, even *Current Contents*. Many of the remaining researchers, who have up to now depended on narrow readings from select journals to meet their needs, may jump on the data base bandwagon while maintaining their narrow outlook unless they accidentally initiate broader profiles which bring them more "fringe" publications. One investigator to whom I have talked pointed out that he preferred a "noisy" or somewhat ill-defined general current awareness profile in many instances since he found it more broadening than a precisely defined profile.

The narrowing tendency of precise current awarensss profiles certainly would not apply, however, to retrospective searches where one is preparing a subject review, planning new directions in a defined research program, or newly exploring an unfamiliar subject area. In these instances, the ISI's *Science Citation Index* and the automated data bases are together the most powerful and now most obvious recourse for any researcher gathering bibliographic information. Here the narrowest definition of the profile subject is required if massive citation inundation is to be prevented. The extreme flexibility of approach implicit here has allowed the investigator to closely tailor his profiles to his unique bibliographic needs. This capacity has probably drawn more and more researchers into data based bibliographic searching. Together, the current awareness and retrospective data base usage must now be increasingly universal among researchers.

Indeed, the impact of this data base usage has been the subject of some concern as expressed by J. K. Martin in a paper which she has submitted recently for publication in *Special Libraries*. According to data presented by Martin, after the initiation of data base services there is a potentially troublesome impact on interlibrary loan activity which can strain a library's resources for a period of some months.

The ARS Current Awareness Literature Service has also created an impact at another point in library services. Some months ago, researchers using the service were warned of a potential library overload in providing Xerox copies and book loans. In the transition quarter last year, USDA researcher requests averaged 450 per day totaling 28,000 for the quarter; a similar (30%) projected increase this year may have severely taxed NAL service capacity.

*Ronald H. Goodwin, Research Entomologist, Insect Pathology Laboratory, Agricultural Research Service, USDA.

Most data bases, unfortunately, do not provide certain critical information which could decrease library impact and overload if incorporated. At present, only the Chemical Abstracts and Commonwealth Agriculture Bureau data bases provide author addresses (as does ISI's *Current Contents*). Many current awareness citations are current enough when provided in print-outs that the researcher could obtain the paper directly from the author if necessary. This service would reduce library impact generally for current papers provided on-site or through interlibrary loans.

Address utilization can be related to abstract coverage. Some data bases provide no abstract and it is often difficult for the researcher to decide on the value of a given paper without seeing it if he is only provided with the paper's title. The Commonwealth Agriculture Bureau Abstracts, for example, are a considerable help in answering the question, "Do I need a copy of this paper right now?" They also increase indirectly the searching power of the profile. Since the title words are expanded by the keywords contained in the printed abstract, the CAB data base must be a very powerful searching system as compared to other data bases. The BIOSIS data bases (Biological Abstracts and Bioresearch Index) seem to provide approximately equal searching power through the use of an extensive indexing system which approximates the CAB Abstracts. In my opinion, the BIOSIS indexing numbers are not as easily readable by the average researcher as are those in the CAB Abstracts and so may not readily provide the researcher with the information that he needs to make his "paper obtaining" decision when faced with an obscure title in his print-out. We should not overlook the significant fraction of relevant papers never obtained when an adequate abstract is included in the data base information. Often, enough information may be provided in the abstract alone for the researcher to incorporate the main findings into his own research problem or to provide the incentive for direct author contact by letter of inquiry.

H. D. Kreilkamp, writing in the July 1977 issue of *College and Research Libraries*, noted that the American Agricultural Economics Data Base had recently been added to the NAL's AGRICOLA data base. In addition to the usual title, author, keyword, abstract, and source code search approaches, this data base also provides the NAL call number. The entire AGRICOLA-CAIN data base system has been enhanced by the inclusion of call numbers throughout ever since the data base was organized; again, library impact has been reduced by the catalogue entry finding time saved, and profile users have been encouraged to seek out relevant publications in the library on their own. On the other hand, Kreilkamp also cited several studies criticizing CAIN's lack of subject category codes and keyword enrichment. Such information could, in the future, be provided through abstracts similar to those used by the Commonwealth Agriculture Bureau. Many researchers retrieve citations on journal titles or Codens; therefore, it would be most useful if the AGRICOLA-CAIN system included the Coden Index for the journal cited since journal title abbreviations vary considerably. Likewise, in order to increase the value of the AGRICOLA-CAIN system to users not present at Beltsville and in anticipation of foreign use of the data base, it would be wise to list the International Standard Serial Number (ISSN) with each citation, in addition to the NAL call number, since call numbers are not uniform among libraries. Such usage would also enhance use of the ISSN's and encourage their further application generally.

The National Agricultural Library's personnel should be proud of the fact that they developed the first major agricultural machine-readable and readily available data base in CAIN. No doubt the subsequently developed data bases have incorporated modifications which have improved upon CAIN's pioneering effort. At present, CAIN may therefore suffer in comparison with them. This was to be expected and merely indicates that CAIN must now be upgraded in order to regain its former respected position.

A comparison was made recently between CAIN and some of the other leading data bases by the USDA Agricultural Research Service. This comparison data will be forthcoming in the journal *Special Libraries* in a paper by H. D. Burton entitled "Multi Data Base Searching in Agriculture: A Cooperative, Computerized Service." Information in the paper was gathered in a semi-controlled manner on a given topic from several user groups. On the basis of both total retrievals and average profile retrieval, the BIOSIS and CAB data bases were equally the most productive of the several data bases under comparison. Because no single data base provides comprehensive coverage, Burton recommended the combination of BIOSIS and CAB as core data bases for agricultural research profiles, to which might be added any of the specialty data bases as the subject required.

I have discussed above some of the probable reasons why CAIN was found to be weaker in retrieval than *Biological Abstracts, Bio-Research Index,* the Commonwealth Agriculture Bureau, and the *Food Science and Technology Abstracts* data bases. Keyword, source code, Coden, and abstract enrichment would increase CAIN's retrieval power. The other suggestions that I have made could enhance the CAIN data base in novel ways, perhaps attracting undecided new data base consumers in competition with other data bases. I understand that the BIOSIS data bases will soon contain author addresses in their citations.

Presently, ARS researchers are free to select any of a number of data bases in preparing profiles for their current awareness programs. The *ARS Current Awareness Literature Service Users Guide* is currently in its third edition (1976). This guide, developed by H. D. Burton, has allowed ARS researchers all over the country to develop their own profiles which are reviewed upon submission and revised in cooperation with the researcher as necessary. The raw novice may at first develop a profile or profiles containing errors of various kinds, some of which probably will not be detected by the CALS review before entry and searching is begun. Corrections, like revisions, are simple enough to enter on the profile plan which is presented as a lead page with each data base search. Valuable time may be lost, however, before the profile error is found and corrected.

Here is an instance where the AGRICOLA on-line service can be most helpful. The terminal operator and the novice researcher (both should be present as a matter of course in using the on-line systems, according to Kreilkamp) together can run the new profile on-line to test its activity against selected data bases and to correct apparent errors. This exercise can also predict the volume of returned material, which can be

critical in either a current awareness or a retrospective search situation. Often the initial profile on a given subject may call down too many or too few citations to meet the precise needs of the researcher. If the profile persistently calls down irrelevant titles after corrections are attempted, the researcher can soon realize the severity of the problem and, after further study, return with a differently slanted attack on the problem subject.

The AGRICOLA on-line service may also be useful in one-time searches of unfamiliar subject areas or when a critical question must be answered quickly. Since only a handful of ARS researchers are at Beltsville, however, any use of on-line data base systems usually must come through laboratory or nearby university facilities at the commercial rate of $40 or more per hour. This cost, plus the requirement that the researcher be present for searches, removes the on-line systems from consideration for regular use by ARS researchers; at present, ARS research budgets which are already minimal cannot fund this kind of activity. Only four other ARS Laboratory sites nationally have access to on-line systems similar to that at the Beltsville station. All of these are near affiliated major libraries which already contain on-line system capacity. The present ARS Current Awareness System costs less than a third of the commercial batch system rate and is available to all ARS researchers and affiliates regardless of location.

In conclusion, let me say that I consider myself extremely fortunate to have regular access to the major data bases. These powerful bibliographic tools have made up-to-date relevant knowledge readily and easily available in depth and breadth as never known before. Such accessibility of relevant knowledge is critical to researchers in a broad-based scattered-source subject field such as comparative or invertebrate pathology.

Scientific research and informational resources
(Courtesy, National Agricultural Library)

THE ADVANTAGES OF SUBFILES WITHIN A CONTINENTAL NETWORKING SYSTEM

by

Gerald R. Ogden*

Information science, inexact at this stage in development, often offers as many contradictions as rewards. Even as we work toward the goal of capturing and disseminating the world's information, it becomes apparent that the objectives we seek become increasingly elusive, if not impossible to achieve. Yet information scientists, as well as others who share no claim to the title, abound with optimism. Plans are proposed to extend bibliographic control over national and international literature output. Studies and research are conducted in efforts to reconcile incompatibilities arising among large numbers of multi-faceted multi-programmed but, most often, singularly supported data bases. We accept the challenge of offering for public access the greater portion of the 54 million plus data base records that are converted into machine readable form.[1] And, technically, we have the energies and capabilities of overcoming many of the problems we now encounter.

Still, technical competence alone will not answer many pressing needs. For while we are pursuing our quest of containing the world's agricultural literature, events occur that cannot be overcome by the man-machine relationship. Expressly I refer to our inability to coordinate and structure a smoothly operating organization within the agricultural library community that can effectively and efficiently, quantitatively and qualitatively establish bibliographical control over published materials. Presently efforts at bibliographic control at the national level are nonsystematic, widely dispersed, decentralized, and disorganized. These conditions arise because documentation is being conducted by individuals and organizations that have established non-homocentric small-scale enterprises with specialized missions, serving limited user groups, and operating within the confines of restricted resources. Manifest in this desperate approach is the generation of a centrifugal force that weakens the bonds of any commonality that might be shared in cooperative efforts to capture and control agricultural literature. It consists principally of inefficiency, especially through duplication, and of restrictions on our ability to govern the whole literature population. This acts as a disservice to user groups.

Abundant reasons are given as to why uncoordinated decentralization has taken place. Time is a factor. The field of information science has not become sufficiently sophisticated to help us achieve our goals. No comprehensive national policy exists to support a coordinated system. Forceful aggressive leadership has been prevented from materializing at the national level. By their abundancy and continuing growth in numbers, small-scale operations virtually defy coordination. General disinterest and disposition to ignore the rationale of such small-scale operations prevail. Why they were formed, how they were formed, their justifications for existence, their missions are seldom questioned.

This may sound as if these operations have come about largely through default but I would argue otherwise. I say that small information-documentation centers exist because of the impracticability for central organizations to fill the many bibliographic dissemination requirements imposed by the entire agricultural research community. Moreover, I contend that the establishment of a cooperative networking system on a continental scale offers the only practical method available for capturing and controlling the total population of agricultural literature in North America. And I also contend that the structure of the network system, in turn, rests on the use of subfiles. Without subfiles, no network can exist. It becomes a matter of expediency for us, then, to act in a positive manner and encourage present and future information-documentation centers to supply subfiles suitable for centralized coordination.

As I have said, information and documentation centers developed and proliferated because of the incapacity of larger centralized units to fill the information services needs of agricultural researchers. In the 1960's libraries, impacted seriously by a burgeoning output of secondary materials, sought means to modernize housekeeping chores, *i.e.*, they wished to bring the various internal and recordkeeping processes under control by introducing standards of uniformity and speed. Computerized systems were introduced. Following the successful introduction of library automation, librarians next proposed, as a by-product of existing automated procedures, the manufacture of data bases containing bibliographic information in printed form.

The introduction of on-line retrieval capabilities, reflecting both the output of cataloging efforts and uniquely identifiable indexing efforts followed soon after the publishing of bibliographies. But diversity characterized these efforts more than compatibility.

During the growth of automated systems design, three general philosophies emerged regarding the generation of computerized bibliographies. The National Agricultural Library (NAL), anxious to be recognized as a world leader in the control and dissemination of agricultural information, chose to centralize its system: that is, the Library assumed full responsibility in the United States for the capture of documents, for their conversion into machine readable form, and for marketing the output. Wide distribution of the information has been achieved through sale of magnetic tapes at low cost. Purchasers, under no copyright laws or other restrictions, have modified and enhanced the tapes to meet specific needs. Users, other than the Library, are thus benefited.

The product, however, is of a primary nature. Tapes contain the most basic of surrogate information that can be used to identify materials. The choice of documents for input is limited, being based for the most part on a group of "core" journals selected by Library personnel. Coverage thus becomes quite discriminate since a heavy emphasis is placed on materials from the physical and life sciences. Finally, the concept of quantity of input, not quality, has been established, based, apparently, on the belief that the worth of a bibliographic data base is maximized by its size.

---

*This paper represents the views of the author and not necessarily those of the Economic Research Service. Gerald R. Ogden, Economic Research Service of the USDA.

The Commonwealth Agricultural Bureaux (CAB) supported by the British government, assumed yet a different posture for the production of bibliographic output. In contrast to the National Agricultural Library's stress on centralization, CAB initiated a decentralized operation with quasi-independent units inputing materials that are either subject- or discipline-oriented. Quality of output is sought as against quantity of product. Expensive and time-consuming tasks of indexing and abstracting are a part of the program: the resulting output of which becomes a useful and aesthetically pleasing series of bibliographies.

Studies indicate that no serious conflict, especially in the duplication of records, developed between the two organizations.[2] It became more a difference in philosophies regarding the best approach to be taken in documenting agricultural literature. Neither system, however, has been without its critics. One commentator states that CAIN, the earlier acronym used for the present AGRICOLA data base produced by NAL, "...has been criticized for being too unselective, and yet not comprehensive enough in its coverage."[3] Another states that the "*Bibliography of Agriculture* is not primarily intended for broad scanning of current literature or for current awareness service, but is a tool of the National Agricultural Library. It has, therefore, several deficiencies if it is [to serve] as a source of quick information on recent literature about any subject."[4]

On the other hand, critics state that the CAB data files are thought to be "too selective and not comprehensive enough in [their] coverage."[5] Observers further claim that "The journals of [the] Commonwealth Agricultural Bureaux were originally intended for agricultural scientists in the Commonwealth countries," although world-wide distribution is experienced. Lastly, it is remarked that "The categories of the literature for different CAB journals is [sic] based on the needs of 30 or 40 years ago in the Colonial Era. The specialization of science and the interdisciplinary nature of much modern research produces difficulties during retrospective searching."[6]

Objections to the CAIN and CAB systems were the basis for a third philosophy. Analysts felt that world-wide needs for the dissemination of agricultural literature, in developed as well as developing countries, have been sufficiently acute to prompt the appointment of AGRIS study teams for the purpose of examining the feasibility of establishing a system that could overcome the objections to both the CAIN and CAB data bases. Briefly stated, the findings, published in 1971, suggested organizing a coordinated combination of interlocking units.(libraries, information centers, documentation centers, clearing houses, and the like) which could share in the maintenance and benefits of a master data bank of information. Study team members reasoned that the existence of so large a body of literature required a division of labor for the maintenance of the data base.

Certainly no organization was prepared to assume the responsibility for forming such a network at the time. As Maltha points out, the term agriculture is very complex and ill-defined in terms of a field. It consists of many disciplines, especially from the physical sciences. More recently, however, increasing emphasis on developing countries and on poverty, education, and health within our own rural areas has attracted researchers from the social sciences. New areas of interest and fields of knowledge have been introduced and absorbed into agricultural research, including suggests Maltha, "food technology, rural sociology, rural planning, landscaping, recreation, and environmental management."[7]

The AGRIS experts were not dismayed. They proceeded to suggest that a networking system might be developed consisting of a decentralized organization built upon a hierarchical structure. At the pinnacle of the formation rested a management center whose functions could include:

1) Providing selected input to the master data file;

2) Financing, either in whole or in part, units of a lower status on the hierarchy;

3) Coordinating and controlling the whole of the system from a technical and organizational standpoint;

4) Acting as the liaison between units;

5) Furnishing magnetic master tapes; and

6) Publishing national bibliographies.

Within the decentralized system, the smaller input centers would:

1) Capture, convert into machine readable form, and input in a predetermined format the surrogate information of especially selected literature;

2) Act as clearing houses for specialized literature;

3) Provide for current awareness and retrospective searches; and

4) Publish special subject- or discipline-oriented bibliographies.

The study groups further recommended that the National Agricultural Library could serve as the management center for the Western Hemisphere; the Commonwealth Agricultural Bureaux could serve for the European continent.[8]

Since 1971, AGRIS, the National Agricultural Library, and the Commonwealth Agricultural Bureaux have extended efforts to initiate and support networking complexes. But, for a multitude of reasons, the greatest of which is the reluctance to cooperate on the part of potential data suppliers, networking grew slowly and unevenly. This is especially true in the United States where no more than a micro-network has developed with the addition of two subfiles to the AGRICOLA data base. Ironically, the formation of this mini-system emerged, not as the result of conscientious efforts to form a unified system, but as the result of user demands within the agricultural community.

Concern over the control of agricultural literature in the United States during the 1960's and 1970's cannot be uniquely associated with the National Agri-

cultural Library. Interest in literature control has also been expressed by various discipline- or mission-oriented groups who became aware that end users or researchers needed emancipation from the problems caused by a proliferating output of published materials. In essence it became a matter of dissatisfaction with existing secondary services and resulted in a seeking out of alternative methods of introducing similar services with a more discipline- or mission-oriented, complete, and timely emphasis.

The growth of secondary services, many commercially supported and practically all patterned along traditional lines of bibliographic control, mushroomed to fill the void. But to the user they, too, presented problems. In the past two decades the number of secondary services expanded at a greater percentage rate than did the literature output over which they were designed to provide bibliographic control.[9] Duplication of titles input remained consistently high despite the fact that the number of journals containing pertinent information expanded. Documentation practices among services often included references to the easily obtainable professional journals and major series publications from government organizations. And researchers and librarians became frustrated over sketchy indexing and classification practices of the published bibliographies.

The effect on the agricultural research community of their inability to easily access needed literature proved onerous to the pursuit of research. Trade-offs between manhours devoted to manual and, even in some instances, machine searching had to be evaluated against research timetables. Frequently tight schedules took precedence although, perhaps, at the risk of the quality of the research. Thoroughly discouraged researchers often abandoned attempts at exhaustive literature searches. Initial retrieval experiences of machine readable records, if unsucessful, often turned users sour. What impact, if any, this has had on the growth of professionalism within the various disciplines has yet to be measured.

The American Agricultural Economics Association (AAEA), like a number of organizations, recognized early the problems arising between the researchers and their literature and resolved to upgrade the existing system. The economists wished to establish a secondary service in the form of a documentation center with the mission of:

1) Capturing and documenting all literature output by agricultural economists in the United States and Canada;

2) Upgrading the documentation of agricultural economics literature by such refinements as the addition of abstracts;

3) Converting the documented materials into machine readable form; and

4) Publishing the machine output as a widely circulated bibliography.

The intentions of the Board of Directors of the Association were credible. The service was designed to heighten the awareness of agricultural economists, enhance research capabilities, and encourage professionalism. To this end the promoters of the secondary service within the AAEA sought end user benefits as the primary reason for the capture and control of information.

Certain characteristics identify themselves with the establishment of small-scale centers, such as the AAEA's American Agricultural Economics Documentation Center (AAEDC). Most notable is the long time lag occurring between the formation of the idea regarding a center and its actual organization. In the case of the AAEDC nearly nine years of dedicated, persistent campaigning passed before the Center was established.[10] Various explanations can be advanced for the delay but, most significantly, it reflects a general disinterest on the part of the organization as a whole to recognize the benefits that accrue from the dissemination of information in bibliographical form within the profession.

The implications arising from the reluctance of the general organization to endorse the initiation of secondary services are serious. For, as in the case of the AAEDC, lack of enthusiasm or non-endorsement affects acquisition policies, if such policies are based on voluntary submission of documents for inputing. The funding of a center may suffer from lack of support. And, to a large extent, the attitudes of potential users determine the degree to which they profit from the service. The above factors, in turn, greatly influence the day-to-day operations of a center for they determine the allocation of resources to the documentation process, to marketing, and to the education of users.

Other challenges in forming and operating centers present themselves as well. It is not common in most disciplines to have a reservoir of knowledge or expertise available among researchers regarding the operation of secondary services. Additionally, there are few instances when authorities are available for ongoing consultation. Traditionally, therefore--and the AAEA is a notable exception to this rule--projects are inaugurated in the absence of feasibility studies, user needs are virtually ignored, and policies are implemented that may be detrimental to the functioning of the center.[11] At some early stage of development, and this applies to secondary services irrespective of size, the operational philosophy of the center assumes the characteristics of its managers. In the absence of a cooperative spirit offered by either a larger coordinating unit or another center, attitudes become introspective. In brief, the small-scale center tends to become a mini-fiefdom--too small to make a significant contribution to the general information gathering structure; too big in the eyes of its supporters to warrant its suspension. Yet expectations run high and serious conflicts arise when a unit's proposed mission is not met.

Through foresight and a cooperative spirit offered by the AAEA and various governmental bodies, the American Agricultural Economics Documentation Center overcame a number of problems encountered by other projects. A valued part of the program was the cooperative agreement between the Center and the National Agricultural Library which allowed the Center's output to be integrated into the CAIN data base as a subfile. It was a mini-network arrangement from which both organizations benefited. The National Agricultural Library could not only pride itself for initiating the network, but also for enhancing the CAIN data base by a subfile of records whose format was extended to include abstracts.

Initially, however, the Documentation Center became the major beneficiary of the cooperative agreement. Through subfiling the major expense and time involved

in writing software routines for a specialized data file was eliminated. A predetermined file structure avoided the need for decision-making on the part of inexperienced Center personnel on the "what, where, and how" of inputing information into a computerized system. The use of a common format, long desired by proponents of networking, also led, when the Center's subfile, AGE, joined CAIN on-line in January 1976, to a single entry point and the use of a single search strategy for accessing the two files. It was through the efforts of the National Agricultural Library, serving as the coordinator, that the smaller Ag Econ file was offered on-line through commercial vendors. And NAL's marketing of magnetic tapes of CAIN data expanded the user audience significantly.

The National Agricultural Library also cooperated in providing document delivery, thus relieving the Center of this responsibility in its early stage of development. Finally, camera-ready copy of AGE data was made available for the production of the short-lived (four years) *Bibliography of Agricultural Economics*. The timely demise of the *Bibliography* resulted from circumstances far removed from the Center's association with CAIN. Yet, hypothetically speaking, if the relationship between the two files had not existed at the time, the Documentation Center might have been abandoned. The publication of the *Bibliography*, after all, was a major part of its mission. Being a subfile of CAIN, however, provided some prestige and the offering of on-line services provided the continuity needed for the Center's maintenance.

Regrettably, being absorbed into the CAIN system did not overcome many persistent problems that Documentation Center personnel faced when, in fact, they might have been resolved had a sophisticated networking program been developed. Low funding restricted management's efforts to hire persons of proper expertise to staff the center. Personnel, therefore, required education and training in indexing, cataloging, the indoctrination of software, hardware, and in general bibliographic procedures. The National Agricultural Library offered no such programs. The Center's director required advice on software development for accessing the system, information on the most efficient means of converting documents to machine readable form, and advice on the leasing of hardware, etc. Again, little help was forthcoming from the Library.

The application of bibliographic information by end users remains the only justification for the documentation process. Marketing thus becomes an essential part of the management of data bases. Broken down into its several components, marketing includes the following: the sale of magnetic tapes; promotion of the data base through demonstration, publications, and the like; continued programs of indoctrination, education, and training of search stragetists *and* end-users. Marketing remains elementary at the national level and yet the Library has supported no well integrated, sophisticated marketing program. The Center, badly needing to pursue such projects, thus has had no foundation from which to work. With no pressing desire for change emanating from either organization, innovation in networking remained repressed for several years.

But do not mistake assessment of past performance for criticism. The smallest of the subfiles within CAIN, by now AGRICOLA, the quiescent attitude taken toward innovative developments by subfile suppliers, and the minimal demands made by suppliers on the Library's resources did not promote change. This does not hold true now, however, nor will demands lessen in the future. The need for agricultural information worldwide is acute. The agricultural research community, now more than ever before, requires that we establish a new increasingly dynamic position in the dissemination of information. Rapid advances in the computer sciences, in communications, and in the theoretical and practical applications of information science provide us with no alternative but to move ahead.

We must understand that networking provides the only means to advance to our destination. Inherently, this, in turn, requires cooperative action and the assimilation of quality subfiles into an integrated system. Further, to achieve our mission takes leadership, the sustenance of which is aggressiveness and innovation. Since the nation, and indeed the world, looks to the National Agricultural Library to act authoritatively in matters of agricultural information dissemination, it must display these qualities if it is to continue its hegemony in the world's agricultural community. And, despite perceived obstacles, the best time for the Library to assume leadership is now.

The lowest common denominators within a network complex are the subfiles for the quantity and quality of these data files determine the value of a system as a whole. From a subfile supplier's point of view, the Library can most easily gain leadership in establishing a North American agricultural information network if it acts aggressively and innovatively in encouraging the establishment of specialized information and documentation centers to create these files. To this end cooperative relationships need to be fostered. Such agreements, however, cannot be based on theoretical approaches. Nor can our national library rest on its prestige or past accomplishments. Rather, it must present practical solid evidence to potential cooperators that they can benefit from sharing resources. Suitable marketing efforts should be undertaken to demonstrate that the Library can assume leadership, be innovative, and has the expertise to lead in cooperative undertakings.

To this end, NAL should establish a special section within the Library whose responsibilities will include marketing, education and training, and research and development in automated retrieval services. Within the overall program, internal functions and policies might well be reviewed and weaknesses overcome. Specifically, increased attention and funding might be given to the automated data processing unit to better allow the group to meet user demands while simultaneously giving the programmers the freedom to become innovative and meet present as well as future needs. To bridge the ever widening chasm in communication, a group consisting of representatives from subfile suppliers should be established with NAL as an active participant. And an active continuing program of research would be needed to seek out ways to better serve community needs.

Long-standing policies within the Library should also be reviewed. It must be recognized, as Maltha does, that the idea of an all-encompassing bibliographical tool for agricultural literature is utopian; the idea must be dismissed that elementary encyclopedic bibliographic identification of documents is sufficient for

research needs. If proper foundations are laid, efficiency and quality can be enhanced by the division of responsibility according to specialization. It is the responsibility of the Library to establish structural supports (underpinnings) by providing leadership through quality control, education, training, and a lending of general expertise to new and established subfile suppliers. NAL cannot hope to finance new centers. It can provide seed money, offer cooperation, actively solicit the participation of government bodies, learned societies, mission- and discipline-oriented associations. It can accomplish much by an early identification of user's needs and, through research, can present alternatives on how to best meet these needs.

The Library simply cannot be all things to all people. As the network coordinator utilizing the division of responsibility concept, it would need to encourage specialized units to supply services such as document delivery, on-line retrospective searching, and publication of subject-oriented literature for specialized groups. Current awareness services, providing these can encompass the needs of the entire agricultural community, should be centralized.

The Library could then implement through cooperative action the marketing for the entire system. Subfile suppliers can best market their own files. The NAL should encourage and support this activity. In addition, the proposed special section to be included in the Library's organization can further advance the program through education and training of information specialists and users. The present seminar activity should be formalized, enriched, and expanded. Both elementary and advanced courses should be offered periodically at the national level. Through the information-documentation centers workshops for users must be initiated. The Library can participate in the activities by developing an educational program, providing materials (outlines, agenda, handouts, etc.), and through participation. Seminars and acceptable forms of promotion can also be undertaken to market the AGRICOLA data file.

The synopsis presented here is not all inclusive in terms of needs to be met in establishing a continental network for agricultural information. Hopefully, it does help establish the fact that only through a network that can draw upon a set of compatible subfiles can we hope to achieve our mission of aiding researchers in their quest for literary materials. Present tendencies in information dissemination within the field of agriculture do not encourage program development using subfiles. The NAL's responsibility, indeed its duty, is to counter this trend. And the Library can make a significant contribution if it will take up the reins of leadership.

REFERENCES

1. Martha E. Williams, "On-Line Problems," *Bulletin of the American Society for Information Science* 3 (April 1977): 14-16.

2. See especially, Charles P. Bourne, *Overlapping Coverage of Bibliography of Agriculture by 15 Other Secondary Services*, Report no. IGC-PA-69-39, Palo Alto, Calif.: Information General Corporation, 1 June 1969, 80 pp; Charles P. Bourne, Madeleine S. Kasson, and Jeanne B. North, *Study and Comparison of the Indexing of the Bibliography of Agriculture in Relation to the Indexing of 15 Other Secondary Services*, Report no. IGC-PA-69-40, Palo Alto, Calif.: Information General Corporation, June 1969, 99 pp.

3. H. J. Mann, "AGRIS, Level 1. The Case for a Centralized Input," In, U. Schutzsack, *Study Group 1: Organizational Aspects*, FAO/DC/AGRIS 2/A, AGRIS Study Team Report (1971), Fiche S1, I 21-I 26. Hereafter cited as Mann, "Centralized Input."

4. D. J. Maltha, "The Relation between Documentation Level 1 and Level 2," In, U. Schutzsack, *Study Group 2: Organizational Aspects*, FAO/DC/AGRIS 2/A, AGRIS Study Team Report (1971), Fiche S1, I 41-I 44. Hereafter cited as Maltha, "Documentation Level 1 and Level 2."

5. Mann, "Centralized Input," p. I 23.

6. Maltha, "Documentation Level 1 and Level 2," p. I 42.

7. D. J. Maltha, "Information Needs in Agriculture," *Quarterly Bulletin of the International Association of Agricultural Librarians and Documentalists* 17, 4 (1972): 170-177.

8. U. Schutzsack, "SG1. Organizational Aspects. Report of the Study Group Leader," In, U. Schutzsack, *Study Group 1: Organizational Aspects*, FAO/DC/AGRIS 2A, AGRIS Study Team Report (1971), Fiche 31, pp. 2-15.

9. Maurice Line, and Stephen Roberts, "The Size, Growth, and Composition of Social Science Literature," *International Social Science Journal* 28 (1), 1976, as cited in Gerald R. Ogden, "Attempts to Control Access to the Literature of Agricultural Economics," *Agricultural Economics Research* 29 (July 1977): 90-95.

10. L. M. Day, "The American Bibliography of Agricultural Economics," *Quarterly Bulletin of the International Association of Agricultural Librarians and Documentalists* 20, 1 (1975): 9-12.

11. A thorough study on the literature of agricultural economics was conducted under the auspices of the American Agricultural Economics Association. See, Isaac Thomas Littleton, "The Bibliographic Organization and Use of the Literature of Agricultural Economics," Ph.D. Dissertation, University of Illinois, December 1967, 212 pp.

Batten, James W., and J. Sullivan Gibson. *Soils, Their Nature, Classes, Distribution, Use and Care*, rev. ed. (Tuscaloosa: The University of Alabama Press, c1970, 1977, 276 pp., $10.00).

This book is designed as a supplementary text for geography courses in secondary schools and colleges. Approximately half of the book is textual material; the rest consists of appendices, references, a glossary, and an index.

The Foreword states that the book "provides a much-needed restatement of the newest ideas about the processes of soil formation, and a simple treatment of soil classification." The book itself, however, presents little that is new and much that is inaccurate. For example:

> The Argid suborder consists of soils that have a white clay horizon;
>
> Latosols are opposite to Podzols;
>
> In general, alluvial soils are not highly productive;
>
> Classes VII and VIII are too erodible and too rough for agricultural use.

The problem is that this is a book about soils written by people who are not soil scientists. There is something wrong about that. A soil scientist knows that most Argids do not have a white clay horizon; "opposite" is hardly the word to use in comparing Latosols to Podzols; many, if not most, alluvial soils are highly productive; not all soils and miscellaneous areas in capability classes VII and VIII are either erodible or rough.

In my opinion, no book on soils published in 1977 should dwell, as this one does, on Calcification, Podzolization, Latozation, and the like as separate processes of soil formation. It was thought that R. W. Simonson, in his "Outline of a Generalized Theory of Soil Genesis" published in *Soil Science Society of America Proceedings* in 1959, had laid those terms to rest for good. All soils have a number of properties in common. Additions, removals, transfers, and transformations of the same constituents take place in horizon differentiation in most if not all soils. Thus, the processes in the formation of Podzols are the same as those in Latosols. It is the relative importance of the various processes that differs from one soil to another.

Much of Chapter 4 is a discussion of the soil classification system published in the 1938 USDA Yearbook *Soils and Men* and later modified. It is presented as one of the classifications now used by USDA soil scientists. The truth is, the 1938 system has not been used officially in USDA for more than 10 years. The authors refer to the system now in use as *Soil Classification: A Comprehensive System, 7th Approximation.* The system was published in 1975 as *Soil Taxonomy.*

There are reasonably good general discussions for the beginning student on the nature, distribution, use, and care of soils. Every now and then, though, a statement such as this one crops up: "Soils of different textures may not have the same pH values." That goes without saying. Soils of similar textures may not either. No correlation between texture and pH is possible.

And there are other statements that tend to furrow the brow. Examples:

> "...leaching (in cold climates) is so pronounced that it gives the soils a distinct acid reaction." (Many cold soils are not acid.)
>
> "...a study of a soil profile in New Jersey showed that the eastern United States has experienced three distinctly different kinds of climate since the last glacial period..." (That is a rather sweeping generalization from the study of one profile.)
>
> "The B horizon is said to be illuviated...." (Not all B horizons are illuviated.)

Then there is the following caption on one of the photographs: "...Sharpsburg silty clay loam series." (Sharpsburg is the name of a series, silty clay loam indicates the texture of the surface layer, Sharpsburg silty clay loam is the name of a soil type.) And at the risk of being castigated as a nitpicker, I must point out that the reader is advised in Chapter I about a certain process to be discussed in Chapter IV. There is no Chapter IV. There is a Chapter 4, but the process is discussed in Chapter 3.

The authors state in the Preface, "Students think more abstractly than ever before and seek to be challenged by concepts rather than by rote memorization of facts." The facts that they do get, however, should be "true" and up-to-date facts, but I'm afraid that this book is less than it should be on both counts. Above all, readers of this book must be selective in what statements they believe; I maintain that is asking too much.

Reviewed by Robert F. Mitchel, USDA Soil Conservation Service.

Chang, K. C., editor. *Food in Chinese Culture: Anthropological and Historical Perspectives.* (New Haven and London: Yale University Press, 1977, 450 pp., $20.00).

This volume covers the history of the multiple uses of food in Chinese culture from the "beginning" (Yangshao culture, ca. 5000-3200 B.C.) to the present. Each chapter, written by a leading anthropologist or historian, covers a major dynastic era. The volume's editor asked the contributors to present the essential facts "with regard to issues looming large in your data or in your mind." Given such a broad mandate, each section tends to be as rich and varied as the Chinese food described.

At one level the volume serves as an encyclopedic handbook of Chinese foods. The available food ingredients, basic diets, the processes of food preparation, including the utensils used, the regional variations in food, and the use of food in ceremonial and ritualistic contexts, both secular and religious, are noted and discussed for each major dynastic period. Occa-

sionally the reader is overwhelmed by long descriptive listings of all grains, vegetables, fruits, nuts, fish, fowl, sweeteners, spices, and beverages available at any given period. While one assumes that the reader as well as the authors tend to heave a sigh of relief at the conclusion of such largely discriptive selections, these sections do serve an important reference-work function.

At a second level the volume covers a wide variety of topics related to the technological, commercial, and economic aspects of food. These include a thorough analysis of the technology of soybean fermentation, delightful descriptions of the widely existing inns and hosteleries for travelers, and of the restaurants and taverns in the market towns and larger cities, a description of the large-scale refrigerated transportation of perishable food products in the Yüan-Ming period, and many references to the very intricate interregional commercial network which provided the specialized food products of wide-ranging geographic areas to many major Chinese cities.

At a third level several of the contributors discuss the social, cosmological, political, and semantic uses of food in Chinese culture. It is a truism that Chinese society was exceptionally well attuned to status and hierarchical considerations and each author provides abundant illustrations of how the countless varieties of food, the amount of food, and the manner of serving it all defined the status and relationship of those involved. Food was invariably more than nutrition. It was a message and one that was clearly understood by every Chinese. The linkage between Chinese food practices and their well-known cosmological beliefs of *yin* and *yang* and the five elements is thoroughly explored as is the use of food as a means of communication with the plethora of spirits. As expressed by one Chinese informant, "We are getting the gods drunk so they will be happy and let our ancestors out of hell." In perhaps the most philosophically oriented discussion of food in Chinese culture, one contributor suggests that the Chinese *attitude* toward food provides basic insights into the primary characteristic of the Chinese "mentality"--in particular, the "nondogmatic" and "rational" character of the Chinese mind."

In his excellent introduction to this volume, the anthropologist, K. C. Chang, set some very high goals and purposes. He expressed the hope that this volume would be the beginning of an effort to "develop a rigorous methodology for the study of food and food eating," that it would be the "beginning of a theoretical and methodological framework for the study of food as a culture rather than as a chemical process. He noted that such a framework was not presently available in anthropological literature. One may conclude, as the reviewer does, that this volume does not provide such a theoretical and methodological framework although it may provide a foundation of data upon which such a framework may be constructed. One may be less sanguine than the editor that "one of the best ways of getting to a culture's heart would be through its stomach." Nevertheless, food, in all of its varied uses, is obviously a vital aspect of every civilization. This volume, while tedious at some points, provides the most scholarly, careful, and solid collection of data and essays available on the subject. Anyone claiming familiarity with any aspect of Chinese civilization, past or present, needs to be familiar with the important new dimension of that civilization that is presented in this volume.

Reviewed by William R. Johnson, Institute for Sino-Soviet Studies, George Washington University.

Galarza, Ernesto. *Farm Workers and Agri-business in California, 1947-1960.* (Notre Dame: University of Notre Dame Press, 1977, xvii, 405 pp., $11.95).

The United Farmworkers recently issued an official declaration ending the nationwide boycott of table grapes and lettuce that had been going on for nearly eight years. Union officials stated that the goals of the boycott had been achieved and that the time had arrived for a reorientation toward building a strong and responsive union for agricultural workers. Cesar Chavez, head of the United Farmworkers, has received a great deal of public attention as the first individual to successfully organize the thousands of migrant agricultural workers in California. And, indeed, while it is true that much of the credit does belong to Chavez who, through exhibiting a dogged determination and adherence to non-violence, succeeded in attracting media coverage and sympathies of large segments of American society, he was not the first to attempt to organize California's agricultural labor force.

Less well known, but superbly recounted in this book, is the fact that agricultural labor unions have been putting roots down in receptive California labor pools for decades. Organizations such as the Southern Tenant Farmers Union attacked the perceived evils of California's labor supply system as early as the 1930's. Labor strikes and walkouts were conducted with varying degrees of success by the National Farm Labor Union and its successor, the National Agricultural Workers Union on representatives of California's agri-business complex from the late 1940's through the 1950's. Galarza, one of the principal organizers of the NFLU and the NAWU, details the events of several of the major strikes and walkouts such as the DiGiorgio strike of 1947 and the Tracy tomato strike of 1950. Each of these actions pales in comparison to the struggle that the union engaged in with corporate agribusiness over the importation of Mexican nationals to work on farms under the Bracero Program. It is Galarza's contention that growers, sensing a need on the part of agricultural workers for domestic stability, perhaps based on union participation, decided that the answer to their problems lay in getting the U.S. Department of Labor to certify that a shortage of labor existed to harvest a crop at a prevailing wage, a wage often set by the growers. The grower, upon certification that such a shortage existed, was then able to contract for the importation of Mexican nationals under a program that was initiated because of the labor shortages caused by World War II. This method of importing Mexican workers continued until 1965. Galarza contends that the success of the Bracero Program resulted, at least partially, in the demise of the National Agricultural Workers Union. Whatever success was gained by the union movement lay in the fact that the seed of worker unrest and discontent had begun to grow.

Much of Galarza's book is autobiographical in nature and serves additionally as an excellent source of information on farm labor unionism. But the larger picture is the one painted with a bold brush by Galarza detailing an entire societal structure controlled by a few agri-business giants. It is a fascinating, if biased, view of agri-business in California.

Reviewed by Tom Fulton, USDA, Economics, Statistics, and Cooperatives Service.

Schor, Joel. *Henry Highland Garnet: A Voice of Black Radicalism in the Nineteenth Century* (Contributions in American History, No. 54. Westport, Conn.: Greenwood Press, 1977, xii, 251 pp., $15.95).

From the time of the American Revolution, Black Americans had worked, as individuals and as members of organizations such as the *Free African Society of Philadelphia*, for the gradual abolition of slavery. However, with the economic success of Eli Whitney's cotton gin and the Missouri Compromise of 1820, Southern participation in and support for the slave system increased. The anti-black riots of 1829 in Cincinnati, Ohio, only illuminated the deplorable position of blacks in American society. In 1830, articulate black leaders met in Philadelphia to share both their ideas and their thinking concerning the problems of blacks in America. By 1833, William Lloyd Garrison, the Tappan brothers and others had founded the *American Anti-Slavery Society* thereby providing the abolitionist movement with a somewhat more formal appearance. Within the abolitionist movement itself could be found varying degrees of intellectual differences in regard to what and how it should be done.

Within this milieu, the biographical study of Henry Highland Garnet portrays the black abolitionist movement as one filled with inner conflict. The author begins with a suspenseful description of Garnet's childhood and formative years in New York City where the threat of racial persecution was a part of daily living. In his growing struggle for the cause of racial freedom, Garnet becomes more assertive in both his thinking and his speaking. His efforts in the area of political action, civil disobedience, and/or militancy and resistance to slavery were at times to the left of and antagonistic to other important abolitionists of his time such as Frederick Douglas.

The biography adds another important documented facet to the still emerging story of black history and culture in the United States. Toward the end of the book and in contrast to the beginning chapters, the reader may find a decline in the interest level. Nevertheless, this biography should be on the shelves of any library maintaining a strong research collection in the area of black history, civil rights, and human rights.

Reviewed by Alan Fusonie, USDA, National Agricultural Library.

Williams, James H., and Doug Murfield, eds. *Agricultural Atlas of Nebraska.* (Lincoln, Nebraska: University of Nebraska Press, 1977, 110 pp., $12.95).

The title indicates that we are to learn about the "science and art of farming; the work of cultivating the soil, producing crops and raising livestock" through the medium of a "book of maps, tables, charts, illustrations, etc." (Webster's definition of agriculture and atlas, respectively.) Furthermore, the introduction suggests that the *Atlas* is:

> More than a statistical documentation of the past and present, it represents an effort to trace the development of the various components of the [agricultural] industry and to provide some insight into the causes and effects of agricultural progress (xv).

Needless to say, this is a rather ambitious task to complete within 110 pages generously sprinkled with pictures. Fortunately, the text does go beyond the mere printing of tables and charts and attempts to interpret, embellish, and elucidate the physical and economic relationships underlying the numbers given in the text. From an academic or textbook viewpoint the book is much too cursory in nature. However, as an overview of Nebraska's agricultural geography and resources, its crop and livestock activities, its agricultural processing and transportation sectors as well as various measures of economic well-being among Nebraska's farm families the book is of considerable value. Due to its parochial subject matter, Nebraskans (and perhaps former Nebraskans) more than other groups will find the book most rewarding.

Chapter 1 discusses Nebraska's various climatic characteristics as well as the quantity, quality, and use of Nebraska's key resources--soil and water. Chapter 2 continues this discussion through such topics as resource development (e.g. irrigation), conservation (e.g. erosion control), and improvement (e.g. fertilizer). Chapter 3 relates geography to variation in selected farm characteristics (e.g. number, type, size, economic class and tenure). Chapter 4 dwells on the geographic and cultural aspects of the large number of crops grown in Nebraska. A concise description of factors affecting the per acre yield of corn over the last 20 years and the historical development of wheat production in the state stand out as examples of topics covered in this chapter. Similarly, geographical and cultural topics related to animal production are given in Chapter 5. Chapter 6 accentuates the economic consequences of the production processes revealed in Chapters 4 and 5. Cost of inputs (i.e. land values, debt financing, labor expenses), gross sales, and farm income combine to indicate the economic well-being of Nebraska's farm families. All of these factors vary by sales class and geographic location within the state as well as by the particular year in which one chooses to measure these indicators (1969 in this case). Chapter 7 traces the movement of farm commodities through the state's agricultural marketing and processing sectors. In conclusion, Chapter 8 endeavors to project various agricultural trends into the future. At least one statement, concerning the possible advent of farm commodity surpluses, has certainly proven apropos.

As an overall impression, I think the text points out quite well the relationship between geography (including climate) and the type, size, and intensity of

input/output activities across the state. Output in the form of different crops and livestock is certainly related to soil type, rainfall, water availability, terrain, etc. Examples of this interdependency include corn acreage in the sandhills which is influenced by water availability, soybeans which require well-drained, fine textured soils found in the eastern parts of the state, and areas such as the West, which are endowed with substantial amounts of grass and forage acreage and require livestock to harvest this crop. Inputs such as fertilizers, chemicals, and machinery are used in varying types and quantities and their use is dictated in part by the climate and soil of each area within the state. As noted in the *Atlas*, much of the research in the state is directed toward adapting crops and livestock to the climate and to geographic areas throughout the state.

This is not to say that geography and climate can account for all the variation in the locational pattern of agricultural commodity production. The reader must remind himself from time to time that agricultural production is influenced by a number of factors including mechanization (e.g. center-pivot sprinklers), federal government farm programs (e.g. oat acreage has been sensitive to farm programs), energy supplies, and agricultural exports. A large number of these factors are not indigenous to Nebraska's farmers; the point being that Nebraska is part of a much larger system. A challenge implicit within the *Atlas* is that the physical scientists (the agronomists, animal scientists, etc.) must continue to find technological and scientific applications which will enable Nebraska's farmers to remain competitive with other agricultural areas of the country and world.

Since soil and water are at the heart of Nebraska's agriculture, it is not surprising that considerable portions of the text focus upon a discussion of the attributes of these resources and the management problems associated with each. The authors might be faulted for their relative neglect of the environmental issues surrounding these resources. A more direct link could have been made between the increased use of agricultural inputs (i.e. fertilizers, chemicals, etc.) and the environmental problems of air, land, and water quality. In this respect, perhaps environmentalists and conservationists are one and the same. Even the issue of land use planning would have been appropriate in the resource discussion.

Perhaps one of the most notable omissions concerns the interaction of the rural and urban sectors of Nebraska's economy. Although a number of statistics are available to describe this interdependency, one such data series could have been included--the number of part-time farmers and farm families with one or more members working off the farm. Obviously, this phenomenon is more prevalent where gross sales are relatively small. A related measure would have been the net income per farm from farm and nonfarm sources by economic size class.

Overall the editors did a commendable job in taking the work of some 25 agricultural specialists and forging a text which is orderly and easy to read. The pictorial and cartographic inserts used freely throughout enhance the appeal of the book.

Reviewed by Stan Daberkow, USDA, Manpower Studies, Economic Research Service.

Selected Publications of Note

Directory of Special Libraries and Information Center 4th edition by Margaret L. Young, Harold C. Young and Anthony T. Kruzas, eds. (Detroit: Gale Research, 1977). Volume 1. 1,187 p. Volume 1 covers nearly 14,000 libraries with cross references, an appendix of networks and consortia, and a 25,731 reference-subject index.

Guide to Dialog Data Bases. (Palo Alto, California: Lockheed, 1977). 800 p. This book contains revised, expanded, and updated information on Dialog files 1 through 50. In depth explanations are given of how t search individual data bases with numerous search samples.

Information Management in the 1980's. Proceedings of the ASIS (American Society for Information Science) Annual Meeting. vol. 14 (Sept. 1977). Compiled by Clayton A. Shepherd. 127 p., 10 microfiche. Among the many important papers presented at this meeting, two may be of special note to our readers: "Zero Base Budgeting at the National Agricultural Library" by David R. Hoyt of the National Agricultural Library, Beltsville, Md., and "The National Library and Information Service Network: A View from the Bottom" by Donald B. Simpson of the Bibliographical Center for Research, Denver, Colorado.

An Introduction to Computer-Based Library Systems by Lucy A. Tedd. (New York: Heyden, 1977) 208 p., 3 appendixes, 2 indexes. This book covers library auto mation of both public and technical services in the early to mid-1970's, giving nearly equal time to user oriented services such as retrospective searching, SD and on-line services as well as the usual treatment and discussion of the mechanization of circulation, acquisitions, and cataloging functions. It attempts to cover developments on both sides of the Atlantic.

The Japanese Flowering Cherry Trees of Washington, D. by Roland M. Jefferson and Alan E. Fusonie. (Washing ton, D.C.: Agricultural Research Service, U.S. Dept. of Agriculture, 1977). 66 p. (National Arboretum Contribution No. 4). Free. Available from the Offic of Government and Public Affairs, U.S.D.A., Washingtc D.C. 20250. Roland Jefferson, botanist with the National Arboretum, and Alan Fusonie, historian with the National Agricultural Library, have put together an excellent history of the advent of the Japanese flowering cherry trees to the Washington area complet with interesting and historical pictures and other documents.

Librarians and Online Services by Pauline Atherton a Roger Christian. (White Plains, N.Y.: Knowledge Industries Publications, Inc., 1977). 100 p. This a study of computer-based reference services, their impact on library staff and patrons, and problems en tailed. The text covers various topics including th extent of on-line services, impact on staff and patrons, financial considerations, modes of operation, marketing, management, and control. In addition to the text, exhibits and appendixes reproduce search forms, user manuals and other aids in use at librari around the country to facilitate comparative studies

Library Networks, 1976-77 by Susan K. Martin. (Whit Plains, N.Y.: Knowledge Industry Publications, [c197 131 p. Appendix, Index, Bibliography. This book di

cusses the scope of networking, the implications of machine-readable data, the Ohio College Library Center (OCLC) and other networks, national efforts, and network management. An appendix describes 24 major networks.

Machine-Readable Social Science Data, a special issue of the *Drexel Library Quarterly* 46 (January 1977) edited by Howard D. White, Assistant Professor at the Graduate School of Library Science, Drexel University. This issue represents the first major library publication on the data archive movement and the role of the data librarian. Copies are available from the *Drexel Library Quarterly*, Graduate School of Library Science, Drexel University, Philadelphia, Pa. 19104.

Our Appalachia. An Oral History edited by Laural Shackelford and Bill Weinberg with photographs by Donald R. Anderson. (New York: Hill and Wang, 1977). This book is an example of what could be done using the resources of the Appalachian Oral History Project briefly described in the previous entry. Memories and experiences weave a rich historical mosaic of mountain life from the beginning of this century through the throes of industrialization to the pressures of today. The interviews reveal the mountain way of life, methods of farming, family life, and social customs, and the effects of changing times.

The Papers of Frederick Law Olmsted, vol. 1 edited by Charles Capen McLaughlin, Charles E. Beveridge, and Victoria Post Ranney. (Baltimore, Md.: Johns Hopkins University Press, 1977). This book is the first in a series of volumes based on the papers of the man who has come to be recognized as the founder of the profession of landscape architecture in America. The volume contains letters and two autobiographical fragments for the years 1822-1852. Olmsted was a perceptive observer of American society. As a gentleman farmer, he advocated scientific techniques in agriculture, while as an observer of the ante-bellum South he argued that planters should prepare their slaves for freedom and citizenship. When he became a landscape architect and city planner, he foresaw how large the American cities would become and designed parks and suburbs that would enhance the lives of their future inhabitants: Central Park in New York, the Fenway in Boston, and the grounds of the U.S. Capitol in Washington. Subsequent volumes will contain Olmsted's letters on the slave-holding South, his civil war correspondence, and also letters and reports from the last thirty years of his life revealing his interest in conservation. The entire series will span the years 1822 to 1903.

Toward a National Library and Information Service Network: The Library Bibliographic Component. (Washington, D.C.: Library of Congress, Network Development Office, 1977). Available from the LC Cataloging Distribution Service. This is a position paper emanating from a series of meetings organized by the Network Development Office of the Library of Congress and funded by the Council on Library Resources to plan the development of the library bibliographic component of the National Library and Information Service Network, a national data base, in cooperation with other network related organizations. It is an authorities methodology report which will be used as a basis for an extensive project to plan an authority control system for the library bibliographic component of the evolving network. It will also be used to study the role of the Library of Congress in the National Library Network which identifies major characteristics of existing or planned networks to determine missing components.

UNIMARC: Universal MARC Format. International Federation of Library Associations and Institutions (IFLA). (April, 1977). This document represents an important step toward realization of Universal Bibliographic Control, where cataloging of a publication is carried out in the country of origin and the resulting cataloging record is made available to other countries through the new international format (UNIMARC).

Union Catalog. (Pippa Passes, Ky.: Appalachian Oral History Project, Alice Lloyd College, 1977). 170 p., 600 annotations. Index. The Appalachian Oral History Project began in 1970 as a research program to collect tape recorded interviews of the history and folklore of the Central Appalachian regions. More than three thousand cassettes have been collected. One fourth of these have been transcribed. Transcripts have also been published in a microfiche edition by Microfilm Corporation of America. There are many agriculturally related entries including cattle raising, hearthside crafts, carding and spinning, cloth making, hog killing and butchering, molasses making, farming, farm life, food and food preservation, gardening, flour mills, herbs, homesteading, land, logging, tobacco, trapping, and many more. Transcripts or tapes listed in the catalog may be ordered by writing to the Campus Director, Appalachian Oral History Project, at the school that conducted the interview. The addresses of the four schools are: Alice Lloyd College, Pippa Passes, Ky. 41844; Appalachian State University, Boone, N.C. 28607; Emory and Henry College, Emory, Va. 24327; and Lee's Junior College, Jackson, Ky. 41339.

USDA Data Base Directory 1977. The directory is a source of information on automated on-line data maintained by all agencies in the U.S. Department of Agriculture. It contains information on each data base including the organization responsible, the Management Information Systems Coordinator, System name, data base name, agency contact, software used and many other data base characteristics. NAL's data base system, AGRICOLA, is included with information listed for the CAIN, FNIC, and AGECON data bases. Available from NAL, Reference Division.

USDA National Agricultural Library. Library Collection Development Policy. (Washington, D. C.: 1977). 44 p. Index. This policy statement was formulated according to guidelines of the American Library Association which were published in *Library Resources and Technical Services* 21 (Winter 1977): 40-47. It reflects the interests of the USDA and the agricultural community at the time of publication and is designed to guide collection building at NAL, to communicate to the users of the library and to the agricultural community the current lines of development of the national collection, and to serve as an example of a collection policy for agricultural libraries.

DIALOG Adds 10 New Data Bases

Of the 10 new data bases added to DIALOG in 1977, the following should prove useful to those in the field of agriculture: Food Science and Technology Abstracts, Library and Information Science Abstracts, and Current Research Information Service (Agriculture). A fourth data base, the British Agricultural Data Base, is now also available on-line through DIALOG, Lockheed Information Systems' computerized information retrieval service. It is the most comprehensive data base in agricultural science, consisting of the Commonwealth Agricultural Bureaux (CAB) Abstracts, produced in England. CAB Abstracts cover every branch of agricultural science, including agricultural economics and policy, animal breeding and genetics, horticulture, soil management, and human nutrition with information gleaned from more than 8,500 scientific journals in 37 languages as well as books, reports, and other publications. The National Agricultural Library has access to DIALOG data bases. For further information and a free catalog of data bases, write to Lockheed Information Systems, Dept. 50-20, 3251 - Hanover, Palo Alto, Calif. 94304.

Lockheed to Add World Textile Abstracts

Shirley Institute has signed an agreement with Lockheed to make World Textile Abstracts available on-line. See Recent and Upcoming Meetings for information on WTA workshop to be held in 1978.

Lockheed to Add Smithsonian Data Base

SSIE Current Projects, produced by the Smithsonian Science Information Exchange of Washington, D.C., will be offered soon. Lockheed's target date is December 1977. The SSIE data base is a valuable source for information on scientific research either in progress or initiated and completed during the most recent two years. It encompasses all fields of basic and applied research in the life, physical, social, and engineering sciences.

SDC Adds New Data Bases

Systems Development Corporation is making available the Derwent General Patents data base, which is the most comprehensive and authoritative patents data base in the world. All areas of technology are covered. Source documents are the 10,000 patents issued weekly by more than 20 leading industrial nations. A distinctive characteristic of this data base is the grouping together of "families" of patents. A family comprises the first published documents ("basic") and, later, issued documents relating to the same invention ("equivalents"). The Patents data base is available on-line only on SDC's ORBIT System.

SDC also added the Food Science and Technology data base to its ORBIT retrieval system. It corresponds in coverage to the printed Food Science and Technology Abstracts and is produced by International Food Information Services. Literature in this file covers all human food commodities and aspects of food processing (except production of raw foods) since January 1969. Special subject areas included are composition, basic food science, microbiology, hygiene, toxicology, economics, standards, legislation, engineering, processing, packaging, and additives. The citations are collected from more than 1,200 journals originating in more than 50 countries, patents from more than 20 countries, and books in all languages. NAL has access to ORBIT data bases.

SDC has purchased the TITUS (Textile Information Treatment Users Service) data base which was developed by the Institut Textile de France with those cooperating in Germany, United Kingdom, Spain, Italy, and Belgium. SDC expects to have the file up on-line in March 1978.

FACTS System Being Developed

FACTS (Fast Agricultural Communication Terminal System) is being developed at Purdue University and, unlike the other two networks, will involve using "smart" terminals which have the capabilities to handle their own programs and calculations. Programs are still being created and written by Purdue faculty, but will cover community development, home economics, 4-H, and agricultural areas. This project is still in the developmental stage, but inquiries can be directed to Rodney Harrington, General Director for FACTS, Agricultural Data Network, Smith Hall, Purdue University, Lafayette, Indiana 47907.

CONSER II

Responsibility for the planning and development of the CONSER II Project was given to the MARC Development Office of the Library of Congress in April 1977. In support of planning for CONSER II, analyses of the MARC serials and of the CONSER "snapshot" data bases was undertaken and programs developed to create files of specified search keys from the MARC serials files and from the CONSER "snapshot" files in order to test the retrieval effectiveness of certain keys and key combinations. The National Agricultural Library and Cornell University libraries as CONSER participants began adding bibliographic records to the CONSER data base in 1977.

Name Change

In 1977 the name of the Federal Library Committee/OCLC program was changed from Federal Libraries Experiment in Cooperative Cataloging (FLECC) to the Federal Library and Information Network (FEDLINK). The National Agricultural Library is a part of FEDLINK.

NAL Adds New Data Base

In 1976, the records of the American Agricultural Economics Documentation Center's data base (AGECON) were added to AGRICOLA, the National Agricultural Library's computerized data files (formerly CAIN), as one of three subunits.

Patricia Condon

The staff of the National Agricultural Library and many friends in the field mourn the death of Patricia A. Condon, Chief, Lending Division. Mrs. Condon died after a four week illness from cancer. A Memorial service was held on March 11 with a very large gathering, a testimonial to her warm personality and friendliness. Mrs. Condon worked with the National Agricultural Library for ten years all in public service areas where she quickly impressed users who came back time and again for her assistance. She will be fondly remembered by her staff and colleagues.

USDA-Reorganization

On January 24, 1978, four USDA agencies -- Agricultural Research Service (ARS), Cooperative State Research Service (CSRS), Extension Service (ES), and the National Agricultural Library (NAL) -- merged to become a new organization, the Science and Education Administration (SEA), U.S. Department of Agriculture.

Index Note

Cumulative indexes for 1976-1977 will be published with the next issue - New Series vol. 3, nos. 1/2 (January/March 1978).

Graduate School Press
U.S. Department of Agriculture

Announces the Publication of

Agricultural Literature:

Proud Heritage--Future Promise

Proceedings of a three-day conference by a group of distinguished scholars

Edited by:

Alan Fusonie and
Leila Moran

An informative and entertaining book focusing on the historical as well as the continual importance of agricultural literature.

371 pages
66 illustrations

$9.95 paperback
$13.50 hardback

A significant contribution to the American Revolution Bicentennial was a symposium highlighting the achievements of 200 years of American agriculture as documented in agricultural literature.

This ambitious project was made possible through the cooperation of the Agricultural History Society; American Association of Nurserymen, Inc.; American Poultry Historical Association; Graduate School, U.S. Department of Agriculture; and Massey - Ferguson, Inc. with the Associates of the National Agricultural Library, Inc.

Historians, book dealers, publishers, editors, folklorists, geographers, librarians, and scientists were invited to participate. They presented papers that reviewed and interpreted sources, histories, and other writings.

Each paper is well documented and the extensive index increases the book's potential as a ready reference tool for students, scholars, and librarians.

". . . there is a modern formalism that must be accepted, if not adopted."

". . . a meticulous guide to Agricultural literature in the United States . . . that should be on the shelves of any library which pretends to supply the fundamentals of agricultural history in whatever country."

(G.E. Fussell, Library Review, 4, 1977)

"A section on the westward movement is especially relevant. . . . Also important to students of the West is a series of essays and comments about documents and periodicals dealing with agriculture and with agricultural libraries."

(The Western Historical Quarterly, October, 1977)

Graduate School Press
U.S. Department of Agriculture
Room 1410, South Building
Washington, D.C. 20250

Agricultural Literature: Proud Heritage--Future Promise

_____ paperback @ $9.95

_____ hardback @ $13.50

☐ Check Enclosed

☐ Purchase Order enclosed*

*We can bill only upon receipt of an official purchase order. Postage and handling will be added.

Name _____

Address _____

City and State _____ Zipcode _____

The Associates of the
National Agricultural Library, Inc.
10301 Baltimore Blvd.
Beltsville MD 20705

THIRD CLASS MAIL

PRINTED MATTER